TRANSACTIONS
OF THE
AMERICAN INSTITUTE
OF
ELECTRICAL ENGINEERS

VOLUME 76
1957

PART I. COMMUNICATION AND ELECTRONICS

PUBLISHED BY THE
AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS
33 WEST 39TH STREET
NEW YORK 18, N. Y., U. S. A.

Copyright 1958 by the American Institute of Electrical Engineers
Printed in U. S. A. by The Mack Printing Company, Easton, Pa.

Preface

The AIEE *Transactions* for 1957 (volume 76) is published in three parts: Part I. Communication and Electronics; Part II. Applications and Industry; and Part III. Power Apparatus and Systems. The papers in each of the three parts are classified according to subject matter as follows:

Part I. Communication and Electronics	**Part II. Applications and Industry**	**Part III. Power Apparatus and Systems**
Communication Switching Systems	Air Transportation	Carrier Current
Communication Theory	Domestic and Commercial Applications	Dielectrics
Radio Communications Systems	Land Transportation	Insulated Conductors
Special Communications Applications	Marine Transportation	Power Generation
Telegraph Systems	Production and Application of Light	Protective Devices
Television and Aural Broadcasting Systems	Chemical Industry	Relays
Wire Communications Systems	Electric Heating	Rotating Machinery
Basic Sciences	Electric Welding	Substations
Computing Devices	Feedback Control Systems	Switchgear
Electrical Techniques in Medicine and Biology	General Industry Applications	System Engineering
Electronics	Industrial Control	Transformers
Instruments and Measurements	Industrial Power Rectifiers	Transmission and Distribution
Magnetic Amplifiers	Industrial Power Systems	
Metallic Rectifiers	Mining and Metal Industry	
Nucleonics	Petroleum Industry	Education
Solid State Devices		Safety
		Research

Each part has been indexed separately in the back of that particular part. The three parts are not cross-referenced; hence the user should determine first whether the subject matter of the paper desired is in the field of communication and electronics, applications and industry, or power apparatus and systems. Papers are listed in the subject index under several key words in the titles. The original numbers assigned to the papers are given in the author index. Volume 76 contains the technical papers and related discussions presented at these meetings:

1. Winter General Meeting, New York, N. Y., January 21–25, 1957.
2. Southern District Meeting, Jackson, Miss., April 3–5, 1957.
3. Great Lakes District Meeting, Des Moines, Iowa, April 15–17, 1957.
4. North Eastern District Meeting, Pittsfield, Mass., May 1–3, 1957.
5. Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957.
6. Pacific General Meeting, Pasco, Wash., August 28–30, 1957.
7. Fall General Meeting, Chicago, Ill., October 7–11, 1957.

Statements and opinions given in the papers and discussions published in *Transactions* are the expressions of the contributors for which the American Institute of Electrical Engineers assumes no responsibility.

Communication and Electronics Index for 1957

1. Technical Subject Index

A-C Amplifiers, Upper Limits of Output Power in Vacuum-Tube and Transistor. Vallese.....87–92
A-C Microammeter, The Clamp-Type. Montgomery, Stansbury........................249–53; disc. 253
A-C Voltage Source, A New Magnetic Amplifier, The D-C Controlled. Manteuffel, McCary.....562–6
Acoustic Ionization Detector. Whithouse, Replogle...406–09
Activation Technique for Reactor-Flux-Profile Measurements, A. Wire-. Klickman, DeFalco.....428–30
Advanced Data System in the Power Industry. Keyser, Leslie..206–10
Air-Borne Control Systems, High-Frequency Magnetic Servo Amplifier for. Seegmiller............526–9
Air Ionization and Its Biological Effects, The Control of. Hicks, Beckett..........................108–11
Alternating Induction, Magnetic Circuits with D-C Magnetization and High. Garbarino.....338–48
Alternators for Parallel Operations, Analog Computer Representation of. Boffi, Haas....153–7; disc. 157
Ammeter, Compensation for Instrument Losses in Circuits Containing a Wattmeter, Voltmeter, and. Vrana.........................325–7; disc. 327
Amplifier for Air-Borne Control Systems, High-Frequency Magnetic Servo. Seegmiller.......526–9
Amplifiers, A Nonlinear Low-Frequency Instability Phenomenon in Audio. Usher..........698–701
Amplifiers for Digital Control, Magnetic Output. Kernick, Perkins...............504–08; disc. 508
Amplifiers, Upper Limits of Output Power in Vacuum-Tube and Transistor A-C. Vallese.........87–92
Amplitudes, Small-Signal Heterodyne Mixers with Excessive Injection. Cline.................739–45
Analog Computer Representation of Alternators for Parallel Operations. Boffi, Haas..153–7; disc. 157
Analog Computer Study of Wind Tunnel Drive. Black, Noorda.........................745–50; disc. 750
Analog Computer, The Transformer. Rao, Krishna...732–8
Analysis and Application Using Block-Diagram Techniques, Magnetic Amplifier. Hubbard....57–766
Analysis in Ladder Networks, The Development of a New Method of Circuit. Coker.158–60; disc. 160
Analysis of Second-Order Flip-Flops, Transient. Vallese.................................161–6
Analysis of the Full-Wave Magnetic Amplifier Circuits Considering the Change of the Width of the Dynamic Hysteresis Loop. Kikuchi..........241–9
Application of the Hurwitz-Routh Stability Criteria, A New. Usher..................530–3; disc. 533
Applications of the Iatron Storage Tube, Characteristic and. Davis.............................47–53
Asymmetrically Connected Sphere Gaps, Factors Influencing the Sparkover Voltage of. Fiegel, Keen..............................307–12; disc. 312
Atomic Power Laboratory, Recent Developments in Nuclear Instrumentation at the Knolls. Stone...396–9
Audio Amplifiers, A Nonlinear Low-Frequency Instability Phenomenon in. Usher...........698–701
Audio Frequencies, Measurement of Voltage Ratio at. Sze..444–9
Automatic Adjusting of Wire Spring Relays, Dynamic Gauging and. Weeks.........................520–6
Automatic Line Concentrator and Subscriber Telephone Applications, The. Evers, Wylie............6–9

B

Balancing Transformer-Core-Loss Bridge, A Self-. Foley...567–73
Ballast Resistor in D-C Power Supplies, Elimination of the. Krakauer...........................101–08
Bandwidth Efficient Binary Transmission System, Kineplex, A. Mosier, Clabaugh..723–7; disc. 727
Bandwidth Requirements, Some Relations Between Television Picture Redundancy and. Powers, Staras.......................492–6; disc. 496
(Basic Sciences) A Describing Function for Multiple Nonlinearities Present in Electrohydraulic Control Valves. Zaborszky, Harrington..........183–90
(Basic Sciences) A New Application of the Hurwitz-Routh Stability Criteria. Usher..530–3; disc. 533
(Basic Sciences) Design Relationships for Iron-Core Filter Chokes. Usher.........................484–7
(Basic Sciences) Electric Nonlinear Computation Performed by Linear Elements. Abou-Hussein...378–80
(Basic Sciences) Equations for the Inductances and Current Distribution of Multiple-Conductor Single-Phase and Polyphase Busses. Chin, Higgins.553–8
(Basic Sciences) Forces and Stresses in an Electromagnetic Field. Lee..............267–71; disc. 271
(Basic Sciences) Generalized Charts of the Effects of Nonlinearities in Electrohydraulic Control Valves. Zaborszky, Harrington...................191–8
(Basic Sciences) Magnetic Circuits with D-C Magnetization and High Alternating Induction. Garbarino...338–48
(Basic Sciences) Solution of Simultaneous Equations by Block Elimination. Lof..................509–14
(Basic Sciences) Temperature Rise of Solid Junctions Under Pulse Load. Diebold..............593–8
(Basic Sciences) The Development of a New Method of Circuit Analysis in Ladder Networks. Coker..........................158–60; disc. 160
Beam Focusing Fields from Transverse Component Determinations, Rectilinearity of Electron-. Cioffi.......................................15–19
Beam Generators, Bias Systems for Resonance Transformer Million-Volt Electron. Westendorp.751–5
Behavior and Magnetic Amplifier Performance, Dynamic Core. Finzi, Critchlow............229–40
Bell System, 6,000-Megacycle-Per-Second Radio Relay System for Broad-Band Long-Haul Service in the. McDavitt..........................715–22
Bell System Speakerphone, The. Clemency, Romanow, Rose..148–53
Bias Circuits, A Generalized Theory of Transistor. Hellerman...................................694–7
Bias Systems for Resonance Transformer Million-Volt Electron Beam Generators. Westendorp...751–5
Binary Transmission System, Kineplex, A Bandwidth Efficient. Mosier, Clabaugh....723–7; disc. 727
Biological Effects, The Control of Air Ionization and Its. Hicks, Beckett..........................108–11
BIZMAC II Tapefile, The. Brustman, Agin.728–31
Block-Diagram Techniques, Magnetic Amplifier Analysis and Application Using. Hubbard......57–766
Block Elimination, Solution of Simultaneous Equations by. Lof..................................509–14
Bonneville Power Administration System, Microwave Radio Communication Under Severe Winter Conditions on the. Strong, Warchol, Benson...655–62
Bridge, A Self-Balancing Transformer-Core-Loss. Foley..567–73
Bridge, An Inductronic Double. Miller.......198–200
Broad-Band Long-Haul Service in the Bell System, 6,000-Megacycle-Per-Second Radio Relay System for. McDavitt.............................715–22
Brookhaven National Laboratory, Safety Circuit Development at. Binns, Lones, Pitcher, Melice.365–8
Brushless Machinery, Silicon Rectifiers in. Henkels, Camp, Fortier, Shilling, Smith.............620–6
BTU Computer: An Advanced Data System for Industry, The. Hannula..........................222–5
Buffers for SAGE Computers, The Elimination of Motor-Generator Sets as Transient. Gano.468–73
Buried Cable Telephone Systems. Chilberg.....130–5
Burn-Out Testing of Nuclear Fuel Elements, Electrical Problems in Electrical. Hunt............400–06
Busses, Equations for the Inductances and Current Distribution of Multiple-Conductor Single-Phase and Polyphase. Chin, Higgins...........553–8

C

Cable Circuits, Multiplex Telegraph Equipment for Radio and Submarine. Diamond.............175–82; disc. 182
Cable Telephone Systems, Buried. Chilberg....130–5
Cables, Loading Coils for Ocean. Newell......478–82
(Canada) Development in Communications in the Lower St. Lawrence Valley. Doak, Tennet..26–39
Canada, Light-Route Microwave Systems in. Dinnin...488–92
Canada, The Development of Television in. Hayes...482–4
Canadian Broadcasting Corporation's Radio Service in Canada, The. Olding..................558–61
Canadian Point of View on Radio-Frequency Spectrum Management, A. Brant...............455–61
Canadian Transcontinental Microwave System, The. Bonneville......................................473–7
Carrier, Mechanical Features of 561 Subscriber. Ross, Goldman...................................677–81
Carrier System, Engineering Features and Field Trial Performance of a New Subscriber. Layburn..681–7
Carrier System for Subscriber Use Employing Transistors, A New 5-Channel. Stachiewicz, Layburn, Jamieson...385–90
Channel Displacement Measuring Device Utilizing Magnetic and Paper Tape Recording, A 14-. Partridge, Dunkin, Anderson, Geldmacher..461–7
Characteristics and Applications of the Iatron Storage Tube. Davis.........................47–53
Charge Imparted to Fine Particles by a Corona Discharge, Measurements of. Penney, Lynch..............294–9; disc. 306
Charging of Small Particles for Electrostatic Precipitation, The. Hewitt..............300–06; disc. 306
Charts of the Effects of Nonlinearities in Electrohydraulic Control Valves, Generalized. Zaborszky, Harrington..191–8
Chokes, Design Relationships for Iron-Core Filter. Usher..484–7
Circuit, A Magnetic Amplifier Voltage-Current Overload Protection. Distler, Sturzenbecker....602–06
Circuit Analysis in Ladder Networks, The Development of a New Method of. Coker.....158–60; disc. 160
Circuit Design, Transistor NOR. Rowe, Royer..263–7
Circuit Development at Brookhaven National Laboratory, Safety. Binns, Lones, Pitcher, Melice..365–8
Circuit for a Magnetic Subharmonic Pulser. Hu, Ku..137–41
Circuit for Semiconductor Rectifiers, A New Voltage-Divider. Dortort...............356–8; disc. 358
Circuit, On the Input Power of the Half-Wave Magnetic Amplifier. Kikuchi, Murakami..........10–14
Circuits, A Generalized Theory of Transistor Bias. Hellerman..694–7
Circuits, A New Family of Transistor Switching. Rubinoff...286–9
Circuits, A Numerical-Graphical Method for Synthesizing Switching. Scheinman...............687–9
Circuits Containing a Wattmeter, Voltmeter, and Ammeter, Compensation for Instrument Losses in. Vrana.........................325–7; disc. 327
Circuits, Discontinuous Transition Time Between Stable States in Ferroresonant. Plotkin..........410–21
Circuits, Multiplex Telegraph Equipment for Radio and Submarine Cable. Diamond.175–82; disc. 182
Circuits, Simultaneous Pulled Oscillations in a Triode Oscillator Incorporating Two Oscillatory. Mostafa..120–7
Circuits with D-C Magnetization and High Alternating Induction, Magnetic. Garbarino.........338–48
Clamp-Type A-C Microammeter, The. Montgomery, Stansbury.....................249–53; disc. 253
Closed-Circuit Network for Educational Television. Warman..440–4
Coding, Moment Detection and. Slade, Nanni, Fich, Molony...275–9
Coils for Ocean Cables, Loading. Newell......478–82
Color Television Transmission, A Modified Microwave System for. McClatchie...................626–33
(Committee Report) Proposed Size Standards for Toroidal Magnetic Tape-Wound Cores (57-206)...135–7
(Communication Switching Systems) A New Family of Transitor Switching Circuits. Rubinoff.....286–9
(Communication Switching Systems) A Numerical-Graphical Method for Synthesizing Switcing Circuits. Scheinman.........................687–9
(Communication Switching Systems) Dynamic Gauging and Automatic Adjusting of Wire Spring Relays. Weeks...520–6
(Communication Switching Systems) Quality Assurance—An Essential Part of the Western Electric Manufacturing Process. Karraker...............540–2
(Communication Switching Systems) The Automatic Line Concentrator and Subscriber Telephone Applications. Evers, Wylie.........................6–9
(Communication Switching Systems) The Examination of Electric Contacts by the Plastic Replica Method. Hermance, Egan.........................756–62
(Communication Switching Systems) Transistor 2-Terminal Switches. Har'el..............328–38
(Communication Theory) Kineplex, A Bandwidth Efficient Binary Transmission System. Mosier, Clabaugh......................723–7; disc. 727
(Communication Theory) Moment Detection and Coding. Slade, Nanni, Fich, Molony..........275–9
(Communication Theory) Some Relations Between Television Picture Redundancy and Bandwidth Requiements. Powers, Staras....492–6; disc. 496

(Communication Theory) What Use Is Delta Modulation to the Transmission Engineer? Bowers. .142–7
Communication Under Severe Winter Conditions on the Bonneville Power Administration System, Microwave Radio. Strong, Warchol, Benson....655–62
Communications in the Lower St. Lawrence Valley, Development in. Doak, Tennet..........26–39
Communications, Information Theory Impact on Modern. Mertz..........431–7
Compensation for Instrument Losses in Circuits Containing a Wattmeter, Voltmeter, and Ammeter. Vrana..........325–7; disc. 327
Complex Dielectric Constant of Very High Dielectric Constant Materials at Microwave Frequencies, Measurement of the. Bady..........225–8
Component Determinations, Rectilinearity of Electron-Beam Focusing Fields from Transverse. Cioffi..........15–19
Computation Performed by Linear Elements, Electric Nonlinear. Abou-Hussein..........378–80
Computer: An Advanced Data System for Industry, The BTU. Hannula..........222–5
Computer, Cycle- and Delay-Time Considerations in a Real-Time Digital. Freeman..........588–93
Computer for a Large-Scale Real-Time Application, The Logical Design of a Digital. Astrahan, Housman, Jacobs, Mayer, Thomas..........71–5
Computer Representation of Alternators for Parallel Operations, Analog. Boffi, Haas. .153–7; disc. 157
Computer Study of Wind Tunnel Drive, Analog. Black, Noorda..........745–50; disc. 750
Computer, The Transformer Analog. Rao, Krishna..........732–8
Computers, Design of Long Passive Oscilloscope Probes to Simplify Maintenance of Large Digital. Santelmann, Hingston..........359–62
Computers, The Elimination of Motor-Generator Sets as Transient Buffers for SAGE. Gano..........468–73
(Computing Devices) Russian to English Machine Translation with Simple Logical Processing. Wall, Niehaus..........709–14
(Computing Devices) Statistical Design Theory for Strictly Digital Sampled Data Systems. Chang..........702–08; disc. 708
(Computing Devices) The BIZMAC II Tapefile. Brustman, Agin..........728–31
(Computing Devices) Transistor NOR Circuit Design. Rowe, Royer..........263–7
Concentrator and Subscriber Telephone Applications, The Automatic Line. Evers, Wylie..........6–9
Concepts in Evaluating Electric Equipment Insulation Systems—The Problems Encountered, New. Manning..........348–55
Conductor Single-Phase and Polyphase Busses, Equations for the Inductances and Current Distribution of Multiple-. Chin, Higgins..........553–8
Considerations in a Real-Time Digital Computer, Cycle- and Delay-Time. Freeman..........588–93
Constant of Very High Dielectric Constant Materials at Microwave Frequencies, Measurement of the Complex Dielectric. Bady..........225–8
Contacts by the Plastic Replica Method, The Examination of Electric. Hermance, Egan..........756–62
Control of Air Ionization and Its Biological Effects, The. Hicks, Beckett..........108–11
Control of Nuclear Power Plants, Dynamics and. Van Rennes, Simons, Gray..........279–85
Control Systems, High-Frequency Magnetic Servo Amplifier for Air-Borne. Seegmiller..........526–9
Control Valves, A Describing Function for Multiple Nonlinearities Present in Electrohydraulic. Zaborszky, Harrington..........183–90
Control Valves, Generalized Charts of the Effects of Nonlinearities in Electrohydraulic. Zaborszky, Harrington..........191–8
Control Windings, Operational Magnetic Amplifiers with Multiple. Kadota, Bourne..........515–20
Core and Rectifier Functions, Dynamic Operation of Magnetic Amplifiers with Real-. Bourne, Nitzan..........646–55
Core Behavior and Magnetic Amplifier Performance, Dynamic. Finzi, Critchlow..........229–40
Core-Loss Bridge, A Self-Balancing Transformer-. Foley..........567–73
Cores, Proposed Size Standards for Toroidal Magnetic Tape-Wound. (Committee Report)..........135–7
Corona Detector, An Electronic Electrometer as a Versatile. Anderson, Kresge..........449–54
Corona Discharge, Measurements of Charge Imparted to Fine Particles by a. Penney, Lynch..........294–9; disc. 306
Counting-Rate Meter and Period Meter, The Logarithmic-Diode. Barrow..........421–7
Criteria, A New Application of the Hurwitz-Routh Stability. Usher..........530–3; disc. 533
Current Distribution of Multiple-Conductor Single-Phase and Polyphase Busses, Equations for the Inductances and. Chin, Higgins..........553–8
Current Overload Protection Circuit, A Magnetic Amplifier Voltage-. Distler, Sturzenbecker. 602–06
Current-Type Self-Balancing Magnetic Amplifier, The Operation of the. Krall..........380–4
Cycle- and Delay-Time Considerations in a Real-Time Digital Computer. Freeman..........588–93

D

Data System for Industry, The BTU Computer: An Advanced. Hannula..........222–5
Data System in the Power Industry, Advanced. Keyser, Leslie..........206–10
D-C Controlled A-C Voltage Source, A New Magnetic Amplifier, The. Manteuffel, McCary..........562–6
D-C Manetization and High Alternating Induction, Magnetic Circuits with. Garbarino..........338–48
D-C Motor Using a Magnetic Amplifier, Speed Control of a. Leonhard..........112–19
D-C Power Supplies, Elimination of the Ballast Resistor in. Krakauer..........101–08
Deficiency of Standards for High-Voltage High-Power Rectifier Transformers. Halloran. 690–2; disc. 692
Delay-Time Considerations in a Real-Time Digital Computer, Cycle- and. Freeman..........588–93
Delta Modulation to the Transmission Engineer, What Use Is. Bowers..........142–7
Demodulators, Magnetically Keyed, Phase-Sensitive. Mark, Johnson, Johannessen..........1–6
Describing Function for Multiple Nonlinearities Present in Electrohydraulic Control Valves, A. Zaborszky, Harrington..........183–90
Design Factors for the Optimization of Multichannel Radio Systems. Parry..........606–20
Design of an Improved 2-Stator Watt-Hour Meter. Smith..........537–40
Design of Long Passive Oscilloscope Probes to Simplify Maintenance of Large Digital Computers. Santelmann, Hingston..........359–62
Design Relationships for Iron-Core Filter Chokes. Usher..........484–7
Design Theory for Strictly Digital Sampled Data Systems, Statistical. Chang..........702–08; disc. 708
Detection and Coding, Moment. Slade, Nanni, Fich, Molony..........275–9
Detector, Acoustic Ionization. Whitehouse, Replogle..........406–09
Detector, An Electronic Electrometer as a Versatile Corona. Anderson, Kresge..........449–54
Detector for Nuclear Reactor Service, A Resistance-Temperature. Sias..........363–5
Development in Communications in the Lower St. Lawrence Valley. Doak, Tennet..........26–39
Development of a New Method of Circuit Analysis in Ladder Networks. Coker..........158–60; disc. 160
Development of a Thermal-Neutron Flux-Measuring Instrument. Weaver, Smith, Chastain..........573–7
Development of Television in Canada, The. Hayes..........482–4
Developments in Nuclear Instrumentation at the Knolls Atomic Power Laboratory, Recent. Stone. .396–9
Device Utilizing Magnetic and Paper Tape Recording, A 14-Channel Displacement Measuring. Partridge, Dunkin, Anderson, Geldmacher..........461–7
Dielectric Constant of Very High Dielectric Constant Materials at Microwave Frequencies, Measurement of the Complex. Bady..........225–8
(Dielectrics) New Concepts in Evaluating Electric Equipment Insulation Systems—The Problems Encountered. Manning..........348–55
Dielectrics, Phenomena Accompanying Transient Low-Voltage Discharges in Liquid. Williams, Smith. Part II..........93–7
Differential-Feedback Type, Self-Balancing Magnetic Amplifiers of the. Geyger..........39–46
Differential Phase-and-Gain Equalizer, A New Video. Clark..........674–6
Digital Computer, Cycle- and Delay-Time Considerations in a Real-Time. Freeman..........588–93
Digital Computer for a Large Scale Real-Time Application, The Logical Design of a. Astrahan, Housman, Jacobs, Mayer, Thomas..........71–5
Digital Computers, Design of Long Passive Oscilloscope Probes to Simplify Maintenance of Large. Santelmann, Hingston..........359–62
Digital Control, Magnetic Output Amplifiers for. Kernick, Perkins..........504–08; disc. 508
Digital Sampled Data Systems, Statistical Design Theory for Strictly. Chang..........702–08; disc. 708
Diode Counting-Rate Meter and Period Meter, The Logarithmic-. Barrow..........421–7
Direct-Acting Electric Recording Instruments, Improved. Clark, Albright..........215–21; disc. 221
Discharge, Measurements of Charge Imparted to Fine Particles by a Corona. Penney, Lynch..........294–9; disc. 306
Discharges in Liquid Dielectrics, Phenomena Accompanying Transient Low-Voltage. Williams, Smith. Part II..........93–7
Discontinuous Transition Time Between Stable States in Ferroresonant Circuits. Plotkin..........410–21
Discriminators, Transistorized Phase. DeSautels. 19–26
Disk Integration, Mass Flowmeter with Retarding. Ballard, Macintyre..........127–30
Distribution of Multiple-Conductor Single-Phase and Polyphase Busses, Equations for the Inductances and Current. Chin, Higgins..........553–8
Diversity in a Microwave Radio System, Some Results with Frequency. Willis..........63–7
Divider Circuit for Semiconductor Rectifiers, A New Voltage-. Dortort..........356–8; disc. 358
Dividers to Steep-Front Impulse Waves, The Response of Resistance Voltage. Rohlfs, Kresge, Fisher..........634–42; disc. 642
Double Amplifiers, Two-Phase Motor Drive Using Static Frequency-. Dick, Biringer, Slemon..........497–503; disc. 503
Double Bridge, An Inductronic. Miller..........198–200
Drive, Analog Computer Study of Wind-Tunnel. Black, Noorda..........745–50; disc. 750
Dynamic Core Behavior and Magnetic Amplifier Performance. Finzi, Critchlow..........229–40
Dynamic Gauging and Automatic Adjusting of Wire Spring Relays. Weeks..........520–6
Dynamic Hysteresis Loop, Analysis of the Full-Wave Magnetic Amplifier Circuits Considering the Change of the Width of the. Kikuchi..........241–9
Dynamic Operation of Magnetic Amplifiers with Real-Core and Rectifier Functions. Bourne, Nitzan..........646–55
Dynamics and Control of Nuclear Power Plants. Van Rennes, Simons, Gray..........279–85
Dynamics of Electronic Self-Balancing Systems. Jacob..........166–74

E

Economics of Multimillion-Joule Inductive Energy Storage. Early, Walker..........320–4; disc. 324
Educational Television, Closed-Circuit Network for. Warman..........440–4
Electric Contacts by the Plastic Replica Method, The Examination of. Hermance, Egan..........756–62
Electric Equipment Insulation Systems—The Problems Encountered, New Concepts in Evaluating. Manning..........348–55
Electric Nonlinear Computation Performed by Linear Elements. Abou-Hussein..........378–80
Electric Recording Instruments, Improved Direct-Acting. Clark, Albright..........215–21; disc. 221
Electrical Problems in Electrical Burn-Out Testing of Nuclear Fuel Elements. Hunt..........400–06
Electrohydraulic Control Valves, A Describing Function for Multiple Nonlinearities Present in. Zaborszky, Harrington..........183–90
Electrohydraulic Control Valves, Generalized Charts of the Effects of Nonlinearities in. Zaborszky, Harrington..........191–8
Electromagnetic Field, Forces and Stresses in an. Lee..........267–71; disc. 271
Electrometer as a Versatile Corona Detector, An Electronic. Anderson, Kresge..........449–54
Electron-Beam Focusing Fields from Transverse Component Determinations, Rectilinearity of. Cioffi..........15–19
Electron Beam Generators, Bias Systems for Resonance Transformer Million-Volt. Westendorp....751–5
Electronic Electrometer as a Versatile Corona Detector, An. Anderson, Kresge..........449–54
Electronic Self-Balancing Systems, Dynamics of. Jacob..........166–74
(Electronics) A Generalized Theory of Transistor Bias Circuits. Hellerman..........694–7
(Electronics) A Nonlinear Low-Frequency Instability Phenomenon in Audio Amplifiers. Usher. .698–701
(Electronics) Characteristics and Applications of the Iatron Storage Tube. Davis..........47–53
(Electronics) Deficiency of Standards for High-Voltage High-Power Rectifier Transformers. Halloran..........690–2; disc. 692
(Electronics) Design of Long Passive Oscilloscope Probes to Simplify Maintenance of Large Digital Computers. Santelmann, Hingston..........359–62
(Electronics) Economics of Multimillion-Joule Inductive Energy Storage. Early, Walker...320–4; disc. 324
(Electronics) Elimination of the Ballast Resistor in D-C Power Supplies. Krakauer..........101–08
(Electronics) Magnetically Keyed, Phase-Sensitive Demodulators. Mark, Johnson, Johannessen. .1–6
(Electronics) Measurements of Charge Imparted to Fine Particles by a Corona Discharge. Penney, Lynch..........294–9; disc. 306
(Electronics) Simultaneous Pulled Oscillations in a Triode Oscillator Incorporating Two Oscillatory Circuits. Mostafa..........120–7
(Electronics) The Charging of Small Particles for Electrostatic Precipitation. Hewitt. .300–06; disc. 306
(Electronics) Transient Analysis of Second-Order Flip-Flops. Vallese..........161–6
(Electronics) Transistorized Phase Discriminators. DeSautels..........19–26
(Electronics) Upper Limits of Output Power in Vacuum-Tube and Transistor A-C Amplifiers. Vallese..........87–92

Electrostatic Precipitation, The Charging of Small Particles for. Hewitt............300–06; disc. 306
Elements, Electrical Problems in Electrical Burn-Out Testing of Nuclear Fuel. Hunt............400–06
Elimination of Motor-Generator Sets as Transient Buffers for SAGE Computers, The. Gano.........468–73
Elimination of the Ballast Resistor in D-C Power Supplies. Krakauer.........................101–08
Energy Storage, Economics of Multimillion-Joule Inductive. Early, Walker.............320–4; disc. 324
Engineer, What Use Is Delta Modulation to the Transmission. Bowers.........................142–7
Engineering Features and Field Trial Performance of a New Subscriber Carrier System. Layburn. .681–7
English Machine Translation with Simple Logical Processing, Russian to. Wall, Niehaus.....709–14
Equalizer, A New Video Differential Phase-and-Gain. Clark.................................674–6
Equations by Block Elimination, Solution of Simultaneous. Lof...............................509–14
Equations for the Inductances and Current Distribution of Multiple-Conductor Single-Phase and Polyphase Busses. Chin, Higgins....................553–8
Equipment for Radio and Submarine Cable Circuits, Multiplex Telegraph. Diamond.175–82; disc. 182
Evaluating Electric Equipment Insulation Systems—The Problems Encountered, New Concepts in. Manning 348–55
Examination of Electric Contacts by the Plastic Replica Method, The. Hermance, Egan..........756–62

F

Factors Influencing the Sparkover Voltage of Asymmetrically Connected Sphere Gaps. Fiegel, Keen ....307–12; disc. 312
Family of Transistor Switching Circuits, A New. Rubinoff.................................286–9
Feedback Type, Self-Balancing Magnetic Amplifiers of the Differential-. Geyger.................39–46
Ferroresonant Circuits, Discontinuous Transition Time Between Stable States in. Plotkin.........410–21
Field, Forces and Stresses in an Electromagnetic. Lee. ....267–71; disc. 271
Field Trial Performance of a New Subscriber Carrier System, Engineering Features and. Layburn.... 681–7
Filter Chokes, Design Relationships for Iron-Core. Usher.................................484–7
Five-Channel Carrier System for Subscriber Use Employing Transistors, A New. Stachiewicz, Layburn, Jamieson..........................385–90
561 Subscriber Carrier, Mechanical Features of. Ross, Goldman..........................677–81
Flip-Flops, Transient Analysis of Second-Order. Vallese.................................161–6
Flowmeter with Retarding Disk Integration, Mass. Ballard, Macintyre......................127–30
Flux Conditions in Radioactive Media with Applications to Radiation Monitoring, Radiation-. Dahlin.... 391–6
Flux-Measuring Instrument, Development of a Thermal-Neutron. Weaver, Smith, Chastain........573–7
Flux Measuring System, Solid-State Neutron-. Gray, Grim, Replogle, Spencer...................368–72
Flux-Profile Measurements, A Wire-Activation Technique for Reactor-. Klickman, DeFalco...428–30
FM Mobiletelephone System, A Narrow-Band Experimental. Miller.........................98–100
FM Systems, Signal-to-Noise Ratios in Strong-Carrier. Urkowitz..........................599–602
Focusing Fields from Transverse Component Determinations, Rectilinearity of Electron-Beam. Cioffi.... 15–19
Forces and Stresses in an Electromagnetic Field. Lee.. ....267–71; disc. 271
14-Channel Displacement Measuring Device Utilizing Magnetic and Paper Tape Recording, A. Partridge, Dunkin, Anderson, Geldmacher.....461–7
Frequencies, Measurement of the Complex Dielectric Constant of Very High Dielectric Constant Materials at Microwave. Bady.................225-8
Frequencies, Measurement of Voltage Ratio at Audio. Sze.................................444–9
Frequency Diversity in a Microwave Radio System, Some Results with. Willis.................63–7
Frequency-Doubler Amplifiers, Two-Phase Motor Drive Using Static. Dick, Biringer, Slemon...... ....497–503; disc. 503
Frequency Multipliers and Their Rating, Magnetic. McMurray:
  Part II. Frequency Sextuplers....289–93; disc. 293
Frequency Spectrum Management, A Canadian Point of View on Radio. Brant...................455–61
Fuel Elements, Electrical Problems in Electrical Burn-Out Testing of Nuclear. Hunt...........400–06
Full-Wave Magnetic Amplifier Circuits Considering the Change of the Width of the Dynamic Hysteresis Loop, Analysis of the. Kikuchi...........241–9
Function for Multiple Nonlinearities Present in Electrohydraulic Control Valves, A Describing. Zaborszky, Harrington.........................183–90

G

Gain Equalizer, A New Video Differential Phase-and-Clark....................................67–6
Gaps, Factors Influencing the Sparkover Voltage of Asymmetrically Connected Sphere. Fiegel, Keen ....307–12; disc. 312
Gauging and Automatic Adjusting of Wire Spring Relays, Dynamic. Weeks...................520–6
General Aspects of Hands-Free Telephony. Emling... 201–05
Generalized Charts of the Effects of Nonlinearities in Electrohydraulic Control Valves. Zaborszky, Harrington.............................191–8
Generalized Theory of Transistor Bias Circuits, A. Hellermann..............................694–7
Generator Sets as Transient Buffers for SAGE Computers, The Elimination of Motor-. Gano..468–73
Generators, Bias Systems for Resonance Transformer Million-Volt Electron Beam. Westendorp...751–5
Graphical Method for Synthesizing Switching Circuits, A Numerical-. Scheinman..................687–9

H

Half-Wave Magnetic Amplifier Circuit, On the Input Power of the. Kikuchi, Murakami.........10–14
Hands-Free Telephone, The. Rodek..........201–14
Hands-Free Telephone, General Aspects of. Emling... 201–05
Heterodyne Mixers with Excessive Injection Amplitudes, Small-Signal. Cline.....................739–45
High-Frequency Magnetic Servo Amplifier for Air-Borne Control Systems. Seegmiller........526–9
High-Voltage High-Power Rectifier Transformers, Deficiency of Standards for. Halloran........... ....690–2; disc. 692
Hurwitz-Routh Stability Criteria, A New Application of the. Usher....................530–3; disc. 533
Hysteresis Loop, Analysis of the Full-Wave Magnetic Amplifier Circuits Considering the Change of the Width of the Dynamic. Kikuchi..........241–9

I

Iatron Storage Tube, Characteristics and Applications of the. Davis.............................47–53
Impact on Modern Communications, Information Theory. Mertz..............................431–7
Improved Direct-Acting Electric Recording Instruments. Clark, Albright................215–21; disc. 221
Improved Single-Phase Watt-Hour Meter, An. Landry, Scamman...............................438–9
Impulse Waves, The Response of Resistance Voltage Dividers to Steep-Front. Rohlfs, Kresge, Fisher.. ....634–42; disc. 642
Inductances and Current Distribution of Multiple-Conductor Single-Phase and Polyphase Busses, Equations for the. Chin, Higgins...............553–8
Induction, Magnetic Circuits with D-C Magnetization and High Alternating. Garbarino.........338–48
Inductive Energy Storage, Economics of Multimillion-Joule. Early, Walker...........320–4; disc. 324
Inductronic Double Bridge, An. Miller.......198–200
Industry, Advanced Data System in the Power. Keyser, Leslie....................................206–10
Industry, the BTU Computer: An Advanced Data System for. Hannula.......................222–5
Information Theory Impact on Modern Communications. Mertz.............................431–7
Injection Amplitudes, Small-Signal Heterodyne Mixers with Excessive. Cline..................739–45
Input Power of the Half-Wave Magnetic Amplifier Circuit, On the. Kikuchi, Murakami........10–14
Instability Phenomenon in Audio Amplifiers, A Nonlinear Low-Frequency. Usher...........698–701
Instrument, Development of a Thermal-Neutron Flux-Measuring. Weaver, Smith, Chastain......573–7
Instrument Losses in Circuits Containing a Wattmeter, Voltmeter, and Ammeter, Compensation for. Vrana..........................325–7; disc. 327
Instrumentation at the Knolls Atomic Power Laboratory, Recent Developments in Nuclear. Stone.... 396–9
(Instruments) A New Voltage Time Recorder for Transient Measurements. Foust, Bhimani........... ....253–61; disc. 261
(Instruments) Design of an Improved 2-Stator Watt-Hour Meter. Smith...................537–40
Instruments, Improved Direct-Acting Electric Recording. Clark, Albright...........215–21; disc. 221
(Instruments) The Clamp-Type A-C Microammeter. Montgomery, Stansbury........249–53; disc. 253
Insulation Systems—The Problems Encountered, New Concepts in Evaluating Electric Equipment. Manning.......................................348–55
Integration, Mass Flowmeter with Retarding Disk. Ballard, Macintyre...........................127–30
International Telecommunication Union in the Field of Telephony, The Work of the. Bloecker, Caldwell. 543–52
Ionization and Its Biological Effects, The Control of Air Hicks, Beckett...........................108–11
Ionization Detector, Acoustic. Whitehouse, Replogle. 406–09
Iron-Core Filter Chokes, Design Relationships for. Usher....................................484–7

J

Junctions Under Pulse Load, Temperature Rise of Solid. Diebold............................593–8

K

(Kelvin) An Inductronic Double Bridge. Miller...... 198–200
Keyed, Phase-Sensitive Demodulators, Magnetically. Mark, Johnson, Johannessen................1–6
Kineplex, A Bandwidth Efficient Binary Transmission System. Mosier, Clabaugh......723–7; disc. 727
Knolls Atomic Power Laboratory, Recent Developments in Nuclear Instrumentation at the. Stone...396–9

L

Ladder Networks, The Development of a New Method of Circuit Analysis in. Coker....158–60; disc. 160
Large-Scale Real-Time Application, The Logical Design of a Digital Computer for a. Astrahan, Housman, Jacobs, Mayer, Thomas..............71–5
Light-Route Microwave Systems in Canada. Dinnin.. 488–92
Line Concentrator and Subscriber Telephone Applications, The Automatic. Evers, Wylie..........6–9
Linear Elements, Electric Nonlinear Computation Performed by. Abou-Hussein..............378–80
Liquid Dielectrics, Phenomena Accompanying Transient Low-Voltage Discharges in. Williams, Smith:
  Part II....................................93–7
Load, Steady-State Operation of the Push-Pull Saturable-Core Transformer with Resistive. Anderson. ....76–86; disc. 86
Load, Temperature Rise of Solid Junctions Under Pulse. Diebold...............................593–8
Loading Coils for Ocean Cables. Newell.......478–82
Logarithmic-Diode Counting-Rate Meter and Period Meter, The. Barrow......................421–7
Logical Design of a Digital Computer for a Large Scale Real-Time Application, The. Astrahan, Housman, Jacobs, Mayer, Thomas......................71–5
Long-Haul Service in the Bell System, 6,000-Megacycle-Per-Second Radio Relay System fpr Broad-Band. McDavitt...................................715–22
Loop, Analysis of the Full-Wave Magnetic Amplifier Circuits Considering the Change of the Width of the Dynamic Hysteresis. Kikuchi..........241–9
Losses in Circuits Containing a Wattmeter, Voltmeter, and Ammeter, Compensation for Instrument. Vrana..........................325–7; disc. 327
Low-Frequency Instability Phenomenon in Audio Amplifiers, A Nonlinear. Usher..........698–701
Low-Voltage Discharges in Liquid Dielectrics, Phenomena Accompanying Transient. Williams, Smith:
  Part II....................................93–7

M

Machine Translation with Simple Logical Processing, Russian to English. Wall, Niehaus........709–14
Machinery, Silicon Rectifiers in Brushless. Henkels, Camp, Fortier, Shilling, Smith...............620–6
Magnetic Amplifier Analysis and Application Using Block-Diagram Techniques. Hubbard....57–766
Magnetic Amplifier Circuit, On the Input Power of the Half-Wave. Kikuchi, Murakami...........10–14
Magnetic Amplifier Circuits Considering the Change of the Width of the Dynamic Hysteresis Loop, Analysis of the Full-Wave. Kikuchi................241–9
Magnetic Amplifier Performance, Dynamic Core Behavior and. Finzi, Critchlow..............229–40
Magnetic Amplifier, Speed Control of a D-C Motor Using a. Leonhard......................112–19
Magnetic Amplifier, The D-C Controlled A-C Voltage Source, A New. Manteuffel, McCary......562–6
Magnetic Amplifier, The Operation of the Current-Type Self-Balancing. Krall................380–4
Magnetic Amplifier Voltage-Current Overload Protection Circuit, A. Distler, Sturzenbecker....602–06
Magnetic Amplifiers of the Differential-Feedback Type, Self-Balancing. Geyger....................39–46
(Magnetic Amplifiers) Steady-State Operation of the Push-Pull Saturable-Core Transformer with Resistive Load. Anderson.............76–86; disc. 86
(Magnetic Amplifiers) Two-Phase Motor Drive Using Static Frequency-Doubler Amplifiers. Dick, Biringer, Slemon.................497–503; disc. 503
Magnetic Amplifiers with Multiple Control Windings, Operational. Kadota, Bourne............515–20

Magnetic Amplifiers with Real-Core and Rectifier Functions, Dynamic Operation of. Bourne, Nitzan 646–55
Magnetic and Paper Tape Recording, A 14-Channel Displacement Measuring Device Utilizing. Partridge, Dunkin, Anderson, Geldmacher 461–7
Magnetic Circuits with D-C Magnetization and High Alternating Induction. Garbarino........338–48
Magnetic Frequency Multipliers and Their Rating. McMurray:
Part II. Frequency Sextuplers.....289–93; disc. 293
Magnetic Output Amplifiers for Digital Control. Kernick, Perkins..........504–08; disc. 508
Magnetic Servo Amplifier for Air-Borne Control Systems, High-Frequency. Seegmiller.........526–9
Magnetic Subharmonic Pulser, Circuit for a. Hu, Ku. 137–41
Magnetic Tape-Wound Cores, Proposed Size Standards for Torodial. (Committee Report).........135–7
Magnetically Keyed, Phase-Sensitive Demodulators. Mark, Johnson, Johannessen..........1–6
Magnetically Regulated Reference Power Supply. Raber..........534–7
Magnetization and High Alternating Induction, Magnetic Circuits with D-C. Garbarino........338–48
Maintenance, A Statistical Approach to Telephone Transmission. Pringle, Gaudet...663–9; disc. 669
Maintenance of Large Digital Computers, Design of Long Passive Oscilloscope Probes to Simplify. Santelmann, Hingston..........359–62
Management, A Canadian Point of View on Radio-Frequency Spectrum. Brant..........455–61
Manufacturing Process, Quality Assurance—An Essential Part of the Western Electric. Karraker..540–2
Mass Flowmeter with Retarding Disk Integration. Ballard, Macintyre..........127–30
Measurement of the Complex Dielectric Constant of Very High Dielectric Constant Materials at Microwave Frequencies. Bady..........225–8
Measurement of Voltage Ratio at Audio Frequencies. Sze..........444–9
Measurements, A New Voltage Time Recorder for Transient. Foust, Bhimani.....253–61; disc. 261
Measurements, A Wire-Activation Technique for Reactor-Flux-Profile. Klickman, DeFalco..428–30
Measurements of Charge Imparted to Fine Particles by a Corona Discharge. Penney, Lynch..........294–9; disc. 306
Measuring Device Utilizing Magnetic and Paper Tape Recording, A 14-Channel Displacement. Partridge, Dunkin, Anderson, Geldmacher.....461–7
Measuring System, Solid-State Neutron-Flux. Gray, Grim, Replogle, Spencer..........368–72
Mechanical Features of 561 Subscriber Carrier. Ross, Goldman..........677–81
Media with Applications to Radiation Monitoring, Radiation-Flux Conditions in Radioactive. Dahlin 391–6
Megacycle-Per-Second Radio Relay System for Broad-Band Long-Haul Service in the Bell System, 6,000-. McDavitt..........715–22
Meter, An Improved Single-Phase Watt-Hour. Landry, Scamman..........438–9
Meter and Period Meter, The Logarithmic-Diode Counting-Rate. Barrow..........421–7
Meter, Design of an Improved 2-Stator Watt-Hour. Smith..........537–40
Method for Synthesizing Switching Circuits, A Numerical-Graphical. Scheinman..........687–9
Method of Circuit Analysis in Ladder Networks, The Development of a New. Coker..158–60; disc. 160
Method, The Examination of Electric Contacts by the Plastic Replica. Hermance, Egan.........756–62
Microammeter, The Clamp-Type A-C. Montgomery, Stansbury..........249–53; disc. 253
Microwave Frequencies, Measurement of the Complex Dielectric Constant of Very High Dielectric Constant Materials at. Bady..........225–8
Microwave Radio Communication Under Severe Winter Conditions on the Bonneville Power Administration System. Strong, Warchol, Benson........655–62
Microwave Radio System, Some Results with Frequency Diversity in a. Willis..........63–7
Microwave Radio Systems, Noise Considerations on Toll Telephone. Combellick, Ferguson.........67–70
Microwave System for Color Television Transmission, A Modified. McClatchie..........626–33
Microwave System, The Canadian Transcontinental. Bonneville..........473–7
Microwave Systems in Canada, Light-Route. Dinnin.. 488–92
Million-Volt Electron Beam Generators, Bias Systems for Resonance Transformer. Westendorp..751–5
Mixers with Excessive Injection Amplitudes, Small-Signal Heterodyne. Cline..........739–45
Mobiletelephone System, A Narrow-Band Experimental FM. Miller..........98–100
Modified Microwave System for Color Television Transmission, A. McClatchie..........626–33
Modulation to the Transmission Engineer, What Use Is Delta. Bowers..........142–7
Moment Detection and Coding. Slade, Nanni, Fich, Molony..........275–9
Monitoring, Radiation-Flux Conditions in Radioactive Media with Applications to Radiation. Dahlin... 391–6
Motor Drive Using Static Frequency-Doubler Amplifiers Two-Phase. Dick, Biringer, Slemon.......... ....497–503; disc. 503
Motor-Generator Sets as Transient Buffers for SAGE Computers, The Elimination of. Gano....468–73
Motor Using a Magnetic Amplifier, Speed Control of a D-C. Leonhard..........112–19
Multichannel Radio Systems, Design Factors for the Optimization of. Parry..........606–20
Multimillion-Joule Inductive Energy Storage, Economics of. Early, Walker..........320–4; disc. 324
Multiple Control Windings, Operational Magnetic Amplifiers with. Kadota, Bourne.........515–20
Multiplex Telegraph Equipment for Radio and Submarine Cable Circuits. Diamond.175–82; disc. 182
Multipliers and Their Rating, Magnetic Frequency. McMurray:
Part II. Frequency Sextuplers.....289–93; disc. 293

N

Narrow-Band Experimental FM Mobiletelephone System, A. Miller..........98–100
Network for Educational Television, Closed-Circuit. Warman..........440–4
Networks, The Development of a New Method of Circuit Analysis in Ladder. Coker..158–60; disc. 160
Neutron Flux-Measuring Instrument, Development of a Thermal-. Weaver, Smith, Chastain.......573–7
Neutron-Flux Measuring System, Solid-State. Gray, Grim, Replogle, Spencer..........368–72
Noise Considerations on Toll Telephone Microwave Radio Systems. Combellick, Ferguson......67–70
Nonlinear Computation Performed by Linear Elements, Electric. Abou-Hussein..........378–80
Nonlinear Low-Frequency Instability Phenomenon in Audio Amplifiers, A. Usher..........698–701
Nonlinearities in Electrohydraulic Control Valves, Generalized Charts of the Effects of. Zaborszky, Harrington..........191–8
Nonlinearities Present in Electrohydraulic Control Valves, A Describing Function for Multiple. Zaborszky, Harrington..........183–90
NOR Circuit Design, Transistor. Rowe, Royer..263–7
Nuclear Fuel Elements, Electrical Problems in Electrical Burn-Out Testing of. Hunt..........400–06
Nuclear Instrumentation at the Knolls Atomic Power Laboratory, Recent Developments in. Stone.396–9
Nuclear Power Plant, Some Operational Problems of the. Jones, Witzke..........373–7
Nuclear Power Plants, Dynamics and Control of. Van Rennes, Simons, Gray..........279–85
Nuclear Power Reactors, Safety Systems for. Jacobs.. 670–3
Nuclear Reactor Service, A Resistance-Temperature Detector for. Sias..........363–5
(Nucleonic Radiation) A Wire-Activation Technique for Reactor-Flux-Profile Measurements. Klickman, DeFalco..........428–30
(Nucleonic Radiation) Acoustic Ionization Detector. Whitehouse, Replogle..........406–09
(Nucleonic Radiation) Radiation-Flux Conditions in Radioactive Media with Applications to Radiation Monitoring. Dahlin..........391–6
(Nucleonic Radiation) Safety Circuit Development at Brookhaven National Laboratory. Binns, Lones, Pitcher, Melice..........365–7
(Nucleonic Radiation) The Logarithmic-Diode Counting-Rate Meter and Period Meter. Barrow..421–7
(Nucleonics) Bias Systems for Resonance Transformer Million-Volt Electron Beam Generators. Westendorp..........751–5
Numerical-Graphical Method for Synthesizing Switching Circuits, A. Scheinman..........687–9

O

Ocean Cables, Loading Coils for. Newell......478–82
On the Input Power of the Half-Wave Magnetic Amplifier Circuit. Kikuchi, Murakami..........10–14
Operation of Magnetic Amplifiers with Real-Core and Rectifier Functions, Dynamic. Bourne, Nitzan... 646–55
Operation of the Current-Type Self-Balancing Magnetic Amplifier, The. Krall..........380–4
Operational Magnetic Amplifiers with Multiple Control Windings. Kadota, Bourne..........515–20
Operational Problems of the Nuclear Power Plant, Some. Jones, Witzke..........373–7
Operations, Analog Computer Representation of Alternators for Parallel. Boffi, Haas...153–7; disc. 157
Optimization of Multichannel Radio Systems, Design Factors for the. Parry..........606–20
Oscillations in a Triode Oscillator Incorporating Two Oscillatory Circuits, Simultaneous Pulled. Mostafa 120–7
Oscilloscope Probes to Simplify Maintenance of Large Digital Computers, Design of Long Passive. Santelmann, Hingston..........359–62
Overload Protection Circuit, A Magnetic Amplifier Voltage-Current. Distler, Sturzenbecker..602–06
Output Power in Vacuum-Tube and Transistor A-C Amplifiers, Upper Limits of. Vallese.......87–92

P

Paper Tape Recording, A 14-Channel Displacement Measuring Device Utilizing Magnetic and. Partridge, Dunkin, Anderson, Geldmacher.....461–7
Parallel Operations, Analog Computer Representation of Alternators for. Boffi, Haas......153–7; disc. 157
Particles by a Corona Discharge, Measurements of Charge Imparted to Fine. Penney, Lynch....... ....294–9; disc. 306
Performance, Dynamic Core Behavior and Magnetic Amplifier. Finzi, Critchlow..........229–40
Performance of a New Subscriber Carrier System, Engineering Features and Field Trial. Layburn..681–7
Period Meter, The Logarithmic-Diode Counting-Rate Meter and. Barrow..........421–7
Phase-and-Gain Equalizer, A New Video Differential. Clark..........674–6
Phase Discriminators, Transistorized. DeSautels..... 19–26
Phase-Sensitive Demodulators, Magnetically Keyed. Mark, Johnson, Johannessen..........1–6
Phase-Shift Telegraph System, A Predicted Wave-Signaling. Heald, Clabaugh....316–19; disc. 319
Phenomena Accompanying Transient Low-Voltage Discharges in Liquid Dielectrics. Williams, Smith:
Part II..........93–7
Phenomenon in Audio Amplifiers, A Nonlinear Low-Frequency Instability. Usher..........698–701
Plant, Some Operational Problems of the Nuclear Power. Jones, Witzke..........373–7
Plants, Dynamics and Control of Nuclear Power. Van Rennes, Simons, Gray..........279–85
Plastic Replica Method, The Examination of Electric Contacts by the. Hermance, Egan........756–62
Polyphase Busses, Equations for the Inductances and Current Distribution of Multiple-Conductor Single-Phase and. Chin, Higgins..........553–8
Power in Vacuum-Tube and Transistor A-C Amplifiers, Upper Limits of Output. Vallese..........87–92
Power Industry, Advanced Data System in the. Keyser, Leslie..........206–10
Power of the Half-Wave Magnetic Amplifier Circuit, On the Input. Kikuchi, Murakami........10–14
Power Plants, Dynamics and Control of Nuclear. Van Rennes, Simons, Gray..........279–85
Power Reactors, Safety Systems for Nuclear. Jacobs.. 670–3
Power Supplies, Elimination of the Ballast Resistor in D-C. Krakauer..........101–08
Power Supply, Magnetically Regulated Reference. Raber..........534–7
Precipitation, The Charging of Small Particles for Electrostatic. Newitt..........300–06; disc. 306
Predicted Wave-Signaling Phase-Shift Telegraph System, A. Heald, Clabaugh......316–19; disc. 319
Problems of the Nuclear Power Plant, Some Operational. Jones, Witzke..........373–7
Processing, Russian to English Machine Translation with Simple Logical. Wall, Niehaus......709–14
Profile Measurements, A Wire-Activation Technique for Reactor-Flux-. Klickman, DeFalco....428–30
Proposed Size Standards for Toroidal Magnetic Tape-Wound Cores. (Committee Report).......135–7
Protection Circuit, A Magnetic Amplifier Voltage-Current Overload. Distler, Sturzenbecker.602–06
Pulled Oscillations in a Triode Oscillator Incorporating Two Oscillatory Circuits, Simultaneous. Mostafa 120–7
Pulse Load, Temperature Rise of Solid Junctions Under. Diebold..........593–8
Pulser, Circuit for a Magnetic Subharmonic. Hu, Ku.. 137–41
Push-Pull Saturable-Core Transformer with Resistive Load, Steady-State Operation of the. Anderson.. ....76–86; disc. 86

Q

Quality Assurance—An Essential Part of the Western Electric Manufacturing Process. Karraker..540–2

R

Radiation-Flux Conditions in Radioactive Media with Applications to Radiation Monitoring. Dahlin... 391–6
Radio and Submarine Cable Circuits, Multiplex Telegraph Equipment for. Diamond.175–82; disc. 182
Radio Communication Under Severe Winter Conditions on the Bonneville Power Administration System, Microwave. Strong, Warchol, Benson.....655–62
(Radio Communications) A Narrow-Band Experimental FM Mobiletelephone System. Miller......98–100

(Radio Communications) Light-Route Microwave Systems in Canada. Dinnin..............488–92
(Radio Communications) Signal-to-Noise Ratios in Strong Carrier FM Systems. Urkowitz...599–602
(Radio Communications) The Canadian Transcontinental Microwave System. Bonneville......473–7
Radio-Frequency Spectrum Management, A Canadian Point of View on. Brant.................455–61
Radio Relay System for Broad-Band Long-Haul Service in the Bell System, 6,000-Megacycle-Per-Second. McDavitt..............................715–22
Radio Service in Canada, The Canadian Broadcasting Corporation's. Olding..................558–61
Radio System, Some Results with Frequency Diversity in a Microwave. Willis.......................63–7
Radio Systems, Design Factors for the Optimization of Multichannel. Parry.......................606–20
Radio Systems, Noise Considerations on Toll Telephone Microwave. Combellick, Ferguson........67–70
Rating, Magnetic Frequency Multipliers and Their. McMurray:
  Part II. Frequency Sextuplers....289–93; disc. 293
Ratio at Audio Frequencies, Measurement of Voltage. Sze.................................444–9
Ratios in Strong-Carrier FM Systems, Signal-to-Noise. Urkowitz.............................599–602
Reactor-Flux-Profile Measurements, A Wire-Activation Technique for. Klickman, DeFalco.......428–30
Reactor Service, A Resistance-Temperature Detector for Nuclear. Sias..........................363–5
Reactors, Safety Systems for Nuclear Power. Jacobs.....................................670–3
Real-Core and Rectifier Functions, Dynamic Operation of Magnetic Amplifiers with. Bourne, Nitzan.....................................646–55
Recent Developments in Nuclear Instrumentation at the Knolls Atomic Power Laboratory. Stone...396–9
Recorder for Transient Measurements, A New Voltage Time. Foust, Bhimani.........253–61; disc. 261
Recording, A 14-Channel Displacement Measuring Device Utilizing Magnetic and Paper Tape. Partridge, Dunkin, Anderson, Geldmacher.....461–7
Recording Instruments, Improved Direct-Acting Electric. Clark, Albright...........215–21; disc. 221
Rectifier Functions, Dynamic Operation of Magnetic Amplifiers with Real-Core, and. Bourne, Nitzan.....................................646–55
Rectifier Transformers, Deficiency of Standards for High-Voltage High-Power. Halloran.............................690–2; disc. 692
Rectifiers, A New Voltage-Divider Circuit for Semiconductor. Dortort.............356–8; disc. 358
Rectifiers in Brushless Machinery, Silicon. Henkels, Camp, Fortier, Shilling, Smith............620–6
Rectilinearity of Electron-Beam Focusing Fields from Transverse Component Determinations. Cioffi.....................................15–19
Redundancy and Bandwidth Requirements, Some Relations Between Television Picture. Powers, Staras.............................492–6; disc. 496
Reference Power Supply, Magnetically Regulated. Raber..............................534–7
Relations Between Television Picture Redundancy and Bandwidth Requirements, Some. Powers, Staras.............................492–6; disc. 496
Relay System for Broad-Band Long-Haul Service in the Bell System, 6,000-Megacycle-Per-Second Radio. McDavitt..............................715–22
Relays, Dynamic Gauging and Automatic Adjusting of Wire Spring. Weeks....................520–6
Replica Method, The Examination of Electric Contacts by the Plastic. Hermance, Egan..........756–62
Representation of Alternators for Parallel Operations, Analog Computer. Boffi, Haas...153–7; disc. 157
Resistance-Temperature Detector for Nuclear Reactor Service, A. Sias..........................363–5
Resistance Voltage Dividers to Steep-Front Impulse Waves, The Response of. Rohlfs, Kresge, Fisher.............................634–42; disc. 642
Resistive Load, Steady-State Operation of the Push-Pull Saturable-Core Transformer with. Anderson.............................76–86; disc. 86
Resistor in D-C Power Supplies, Elimination of the Ballast. Krakauer..........................101–08
Resonance Transformer Million-Volt Electron Beam Generators, Bias Systems for. Westendorp..751–5
Response of Resistance Voltage Dividers to Steep-Front Impulse Waves, The. Rohlfs, Kresge, Fisher.............................634–42; disc. 642
Results with Frequency Diversity in a Microwave Radio System, Some. Willis.....................63–7
Retarding Disk Integration, Mass Flowmeter with. Ballard, Macintyre........................127–30
Russian to English Machine Translation with Simple Logical Processing. Wall, Niehaus........709–14

S

Safety Circuit Development at Brookhaven National Laboratory. Binns, Lones, Pitcher, Melice..365–8
Safety Systems for Nuclear Power Reactors. Jacobs.....................................670–3
SAGE Computers, The Elimination of Motor-Generator Sets as Transient Buffers for. Gano.......468–73
St. Lawrence Valley, Development in Communications in the Lower. Doak, Tennet..............26–39
Sampled Data Systems, Statistical Design Theory for Strictly Digital. Chang........702–08; disc. 708
Saturable-Core Transformer with Resistive Load, Steady-State Operation of the Push-Pull. Anderson.............................76–86; disc. 86
Second-Order Flip-Flops, Transient Analysis of. Vallese..................................161–6
Self-Balancing Magnetic Amplifier, The Operation of the Current-Type. Krall..................380–4
Self-Balancing Magnetic Amplifiers of the Differential-Feedback Type. Geyger.................39–46
Self-Balancing Systems, Dynamics of Electronic. Jacob.....................................166–74
Self-Balancing Transformer-Core-Loss Bridge, A. Foley.....................................567–73
Semiconductor Rectifiers, A New Voltage-Divider Circuit for. Dortort..............356–8; disc. 358
Servo Amplifier for Air-Borne Control Systems, High-Frequency Magnetic. Seegmiller..........526–9
Signal Heterodyne Mixers with Excessive Injection Amplitudes, Small-. Cline................739–45
Signal-to-Noise Ratios in Strong Carrier FM Systems. Urkowitz.............................599–602
Signaling Phase-Shift Telegraph System, A Predicted Wave-. Heald, Clabaugh.......316–19; disc. 319
Silicon Rectifiers in Brushless Machinery. Henkels, Camp, Fortier, Shilling, Smith............620–6
Simultaneous Equations by Block Elimination, Solution of. Lof............................509–14
Simultaneous Pulled Oscillations in a Triode Oscillator Incorporating Two Oscillatory Circuits. Mostafa.....................................120–7
Single-Phase and Polyphase Busses, Equations for the Inductances and Current Distribution of Multiple-Conductor. Chin, Higgins.................553–8
Single-Phase Watt-Hour Meter, An Improved. Landry, Scamman..............................438–9
6,000-Megacycle-Per-Second Radio Relay System for Broad-Band Long-Hall Service in the Bell System. McDavitt..............................715–22
Size Standards for Toroidal Magnetic Tape-Wound Cores, Proposed. (Committee Report).....135–7
Small-Signal Heterodyne Mixers with Excessive Injection Amplitudes. Cline.................739–45
Solid Junctions Under Pulse Load, Temperature Rise of. Diebold..............................593–8
(Solid State Devices) Discontinuous Transition Time Between Stable States in Ferroresonant Circuits. Plotkin..............................410–21
Solid-State Neutron-Flux Measuring System. Gray, Grim, Replogle, Spencer...............368–72
Some Results with Frequency Diversity in a Microwave Radio System. Willis.....................63–7
Sparkover Voltage of Asymmetrically Connected Sphere Gaps, Factors Influencing the. Fiegel, Keen.............................307–12: disc. 312
Speakerphone, The Bell System. Clemency, Romanow, Rose..............................148–53
Spectrum Management, A Canadian Point of View on Radio-Frequency. Brant..................455–61
Speed Control of a D-C Motor Using a Magnetic Amplifier. Leonhard.......................112–19
Sphere Gaps, Factors Influencing the Sparkover Voltage of Asymmetrically Connected. Fiegel, Keen.............................307–12; disc. 312
Spring Relays, Dynamic Gauging and Automatic Adjusting of Wire. Weeks.....................520–6
Stability Criteria, A New Application of the Hurwitz-Routh. Usher..................530–3; disc. 533
Standards for High-Voltage High-Power Rectifier Transformers, Deficiency of. Halloran..690–2; disc. 692
Standards for Toroidal Magnetic Tape-Wound Cores, Proposed Size. (Committee Report)........135–7
Static Frequency-Doubler Amplifiers, Two-Phase Motor Drive Using. Dick, Biringer, Slemon.............................497–503; disc. 503
Statistical Approach to Telephone Transmission Maintenance, A. Pringle, Gaudet.....663–9; disc. 669
Statistical Design Theory for Strictly Digital Sampled Data Systems. Chang..........702–08; disc. 708
Steady-State Operation of the Push-Pull Saturable-Core Transformer with Resistive Load. Anderson.............................76–86; disc. 86
Steep-Front Impulse Waves, The Response of Resistance Voltage Dividers to. Rohlfs, Kresge, Fisher.............................634–42; disc. 642
Storage, Economics of Multimillion-Joule Inductive Energy. Early, Walker..........320–4; disc. 324
Storage Tube, Characteristics and Applications of the Iatron. Davis.......................47–53
Stresses in an Electromagnetic Field, Forces and. Lee...........................267–71; disc. 271
Strong-Carrier FM Systems, Signal-to-Noise Ratios in. Urkowitz..........................599–602
Study of Wind Tunnel Drive, Analog Computer. Black, Noorda................745–50; disc. 750
Subharmonic Pulser, Circuit for a Magnetic. Hu, Ku.....................................137–41
Submarine Cable Circuits, Multiplex Telegraph Equipment for Radio and. Diamond..175–82; disc. 182
Subscriber Carrier, Mechanical Features of 561. Ross, Goldman..............................677–81
Subscriber Carrier System, Engineering Features and Field Trial Performance of a New. Layburn.681–7
Subscriber Telephone Applications, The Automatic Line Concentrator and. Evers, Wylie........6–9
Subscriber Use Employing Transistors, A New 5-Channel Carrier System for. Stachiewicz, Layburn, Jamieson..............................385–90
Switches, Transistor 2-Terminal. Har'el.......328–38
Switching Circuits, A New Family of Transistor. Rubinoff..............................286–9
Switching Circuits, A Numerical-Graphical Method for Synthesizing. Scheinman.................687–9
Synthesizing Switching Circuits, A Numerical-Graphical Method for. Scheinman.................687–9
System in the Power Industry, Advanced Data. Keyser, Leslie..............................206–10
Systems, Buried Cable Telephone. Chilberg.....130–5
Systems, Dynamics of Electronic Self-Balancing. Jacob.....................................166–74

T

Tape Recording, A 14-Channel Displacement Measuring Device Utilzing Magnetic and Paper. Partridge, Dunkin, Anderson, Geldmacher............461–7
Tape-Wound Cores, Proposed Size Standards for Toroidal Magnetic. (Committee Report)...135–7
Tapefile, The BIZMAC II. Brustman, Agin...728–31
Technique for Reactor-Flux-Profile Measurements, A Wire-Activation. Klickman, DeFalco.....428–30
Techniques, Magnetic Amplifier Analysis and Application Using Block-Diagram. Hubbard...57–766
Telecommunication Union in the Field of Telephony, The Work of the International. Bloecker, Caldwell.....................................543–52
Telegraph Equipment for Radio and Submarine Cable Circuits, Multiplex. Diamond..175–82; disc. 182
Telegraph System, A Predicted Wave-Signaling Phase-Shift. Heald, Clabaugh........316–19; disc. 319
(Telegraph Systems) Loading Coils for Ocean Cables. Newell..............................478–82
Telephone Applications, The Automatic Line Concentrator and Subscriber. Evers, Wylie..........6–9
Telephone Microwave Radio Systems, Noise Considerations on Toll. Combellick, Ferguson........67–70
Telephone System, A Narrow-Band Experimental FM Mobile-. Miller.........................98–100
Telephone Systems, Buried Cable. Chilberg....130–5
(Telephone) The Bell System Speakerphone. Clemency, Romanow, Rose.......................148–53
Telephone, The Hands-Free. Rodek..........210–14
Telephone Transmission Maintenance, A Statistical Approach to. Pringle, Gaudet...663–9; disc. 669
Telephone Transmission, Where We Are and Where We Are Going in. Huntley................54–63
Telephone, General Aspects of Hands-Free. Emling.....................................201–05
Telephony, The Work of the International Telecommunication Union in the Field of. Bloecker, Caldwell..............................543–52
(Television) A New Video Differential Phase-and-Gain Equalizer. Clark..........................674–6
Television, Closed-Circuit Network for Educational. Warman..............................440–4
Television in Canada, The Development of. Hayes.....................................482–4
Television Picture Redundancy and Bandwidth Requirements, Some Relations Between. Powers, Staras..............492–6; disc. 496
(Television) The Canadian Broadcasting Corporation's Radio Service in Canada. Olding........558–61
Television Transmission, A Modified Microwave System for Color. McClatchie.................626–33
Temperature Detector for Nuclear Reactor Service, A Resistance-. Sias..........................363–5
Temperature Rise of Solid Junctions Under Pulse Load. Diebold..............................593–8
Testing of Nuclear Fuel Elements, Electrical Problems in Electrical Burn-Out. Hunt.............400–06
Theory for Strictly Digital Sampled Data Systems, Statistical Design. Chang.........702–08; disc. 708
Theory Impact on Modern Communications, Information. Mertz..............................431–7
Theory of Transistor Bias Circuits, A Generalized. Hellerman..............................694–7
Thermal-Neutron Flux-Measuring Instrument, Development of a. Weaver, Smith, Chastain...573–7
Time Between Stable States in Ferroresonant Circuits, Discontinuous Transition. Plotkin........410–21
Time Digital Computer, Cycle- and Delay-Time Considerations in a Real-. Freeman..........588–93
Time Recorder for Transient Measurements, A New Voltage. Foust, Bhimani.......253–61; disc. 261
Toll Telephone Microwave Radio Systems, Noise Considerations on. Combellick, Ferguson.......67–70
Toroidal Magnetic Tape-Wound Cores, Proposed Size Standards for. (Committee Report)........135–7

Transcontinental Microwave System, The Canadian. Bonneville....473–7
Transformer Analog Computer, The. Rao, Krishna....732–8
Transformer-Core-Loss Bridge, A Self-Balancing. Foley....567–73
Transformer Million-Volt Electron Beam Generators, Bias Systems for Resonance. Vestendorp....751–5
Transformer with Resistive Load, Steady-State Operation of the Push-Pull Saturable-Core. Anderson....76–86; disc. 86
Transformers, Deficiency of Standards for High-Voltage High-Power Rectifier. Halloran....690–2; disc. 692
Transient Analysis of Second-Order Flip-Flops. Vallese....161–6
Transient Buffers for SAGE Computers, The Elimination of Motor-Generator Sets as. Gano....468–73
Transient Low-Voltage Discharges in Liquid Dielectrics, Phemomena Accompanying. Williams, Smith. Part II....93–7
Transient Measurements, A New Voltage Time Recorder for. Foust, Bhimani....253–61; disc. 261
Transistor A-C Amplifiers, Upper Limits of Output Power in Vacuum-Tube and. Vallese....87–92
Transistor Bias Circuits, A Generalized Theory of. Hellerman....694–7
Transistor NOR Circuit Design. Rowe, Royer....263–7
Transistor Switching Circuits, A New Family of. Rubinoff....286–9
Transistor 2-Terminal Switches. Har'el....328–38
Transistorized Phase Discriminators. DeSautels....19–26
Transistors, A New 5-Channel Carrier System for Subscriber Use Employing. Stachiewicz, Layburn, Jamieson....385–90
Transition Time Between Stable States in Ferroresonant Circuits, Discontinuous. Plotkin....410–21
Translation with Simple Logical Processing, Russian to English Machine. Wall, Niehaus....709–14
Transmission, A Modified Microwave System for Color Television. McClatchie....626–33
Transmission Engineer, What Use Is Delta Modulation to the. Bowers....142–7
Transmission Maintenance, A Statistical Approach to Telephone. Pringle, Gaudet....663–9; disc. 669
Transmission System, Kineplex, A Bandwidth Efficient Binary. Mosier, Clabaugh....723–7; disc. 727
Transmission, Where We Are and Where We Are Going in Telephone. Huntley....54–63
Transverse Component Determinations, Rectilinearity of Electron-Beam Focusing Fields from. Cioffi....15–19
Triode Oscillator Incorporating Two Oscillatory Circuits, Simultaneous Pulled Oscillations in a. Mostafa....120–7
Tube, Characteristics and Applications of the Iatron Storage. Davis....47–53
Two-Phase Motor Drive Using Static Frequency-Doubler Amplifiers. Dick, Biringer, Slemon....497–503; disc. 503
2-Stator Watt-Hour Meter, Design of an Improved. Smith....537–40
Two-Terminal Switches, Transistor. Har'el....328–38

U

Union in the Field of Telephony, The Work of the International Telecommunication. Bloecker, Caldwell....543–52
Upper Limits of Output Power in Vacuum-Tube and Transistor A-C Amplifiers. Vallese....87–92

V

Vacuum-Tube and Transistor A-C Amplifiers, Upper Limits of Output Power in. Vallese....87–92
Valves, A Describing Function for Multiple Nonlinearities Present in Electrohydraulic Control. Zaborszky, Harrington....183–90
Valves, Generalized Charts of the Effects of Nonlinearities in Electrohydraulic Control. Zaborszky, Harrington....191–8
Video Differential Phase-and-Gain Equalizer, A New. Clark....674–6
Voltage-Current Overload Protection Circuit, A Magnetic Amplifier. Distler, Sturzenbecker....602–06
Voltage-Divider Circuit for Semiconductor Rectifiers, A New. Dortort....356–8; disc. 358
Voltage Dividers to Steep-Front Impulse Waves, The Response of Resistance. Rohlfs, Kresge, Fisher....634–42; disc. 642
Voltage of Asymmetrically Connected Sphere Gaps, Factors Influencing the Sparkover. Fiegel, Keen....307–12; disc. 312
Voltage Ratio at Audio Frequencies, Measurement of. Sze....444–9
Voltage Source, A New Magnetic Amplifier, The D-C Controlled A-C. Manteuffel, McCary....562–6
Voltage Time Recorder for Transient Measurements, A New. Foust, Bhimani....253–61; disc. 261
Voltmeter, and Ammeter, Compensation for Instrument Losses in Circuits Containing a Wattmeter. Vrana....325–7; disc. 327

W

Watt-Hour Meter, An Improved Single-Phase. Landry, Scamman....438–9
Watt-Hour Meter, Design of an Improved 2-Stator. Smith....537–40
Wattmeter, Voltmeter, and Ammeter, Compensation for Instrument Losses in Circuits Containing a. Vrana....325–7; disc. 327
Wave-Signaling Phase-Shift Telegraph System, A Predicted. Heald, Clabaugh....316–19; disc. 319
Waves, The Response of Resistance Voltage Dividers to Steep-Front Impulse. Rohlfs, Kresge, Fisher....634–42; disc. 642
Western Electric Manufacturing Process, Quality Assurance—An Essential Part of the. Karraker....540–2
What Use is Delta Modulation to the Transmission Engineer? Bowers....142–7
What We Are and Where We Are Going in Telephone Transmission. Huntley....54–63
Wind Tunnel Drive, Analog Computer Study of. Black, Noorda....745–50; disc. 750
Windings, Operational Magnetic Amplifiers with Multiple Control. Kadota, Bourne....515–20
Winter Conditions on the Bonneville Power Administration System, Microwave Radio Communication Under Severe. Strong, Warchol, Benson....655–62
Wire-Activation Technique for Reactor-Flux-Profile Measurements, A. Klickman, DeFalco....428–30
(Wire Communications) A New 5-Channel Carrier System for Subscriber Use Employing Transistors. Stachiewicz, Layburn, Jamieson....385–90
(Wire Communications) A Statistical Approach to Telephone Transmission Maintenance. Pringle, Gaudet....663–9; disc. 669
(Wire Communications) Buried Cable Telephone Systems. Chilberg....130–5
(Wire Communications) Engineering Features and Field Trial Performance of a New Subscriber Carrier System. Layburn....681–7
(Wire Communications) General Aspects of Hands-Free Telephony. Emling....201–05
(Wire Communications) Mechanical Features of 561 Subscriber Carrier. Ross, Goldman....677–81
(Wire Communications) The Bell System Speakerphone. Clemency, Romanow, Rose....148–53
(Wire Communications) The Hands-Free Telephone. Rodek....210–14
(Wire Communications) The Work of the International Telecommunication Union in the Field of Telephony. Bloecker, Caldwell....543–52
(Wire Communications) Where We Are and Where We Are Going in Telephone Transmission. Huntley....54–63
Wire Spring Relays, Dynamic Gauging and Automatic Adjusting of. Weeks....520–6
Work of the International Telecommunication Union in the Field of Telephony, The. Bloecker, Caldwell....543–52

2. Author Index

Abou-Hussein, M. S. M. Electric Nonlinear Computation Performed by Linear Elements. (57-746)....378–80
Agin, J.; J. A. Brustman. The BIZMAC II Tapefile. (57-674)....728–31
Albright, H. E.; E. W. Clark. Improved Direct-Acting Electric Recording Instruments. (57-220)....215–21; disc. 221
Anderson, J. G.; J. S. Kresge. An Electronic Electrometer as a Versatile Corona Detector. (57-632)....449–54
Anderson, R. E. Steady-State Operation of the Push-Pull Saturable-Core Transformer with Resistive Load. (57-96)....76–86; disc. 86
Anderson, R. L.; G. R. Partridge, J. W. Dunkin, R. C. Geldmacher. A 14-Channel Displacement Measuring Device Utilizing Magnetic and Paper Tape Recording. (57-662)....461–7
Astrahan, M. M.; B. Housman, J. F. Jacobs, R. P. Mayer, W. H. Thomas. The Logical Design of a Digital Computer for a Large Scale Real-Time Application. (57-200)....71–5

B

Bady, Isidore. Measurement of the Complex Dielectric Constant of Very High Dielectric Constant Materials at Microwave Frequencies. (56-260)....225–8
Ballard, R. G.; J. T. Macintyre. Mass Flowmeter with Retarding Disk Integration. (57-113)....127–30
Barrow, Bruce B. The Logarithmic-Diode Counting-Rate Meter and Period Meter. (57-403)....412–7
Beckett, John C.; W. W. Hicks. The Control of Air Ionization and Its Biological Effects. (57-73)....108–11
Benson, S. E.; E. M. Strong, E. J. Warchol. Microwave Radio Communication Under Severe Winter Conditions on the Bonneville Power Administration System. (57-915)....655–62
Bergen, Arthur D. Disc....708
Bhimani, B. V.; C. M. Foust. A New Voltage Time Recorder for Transient Measurements. (57-214)....253–61; disc. 262
Bills, G. W. Disc....157, 750
Binns, J. E.; W. Lones, D. G. Pitcher, M. Melice. Safety Circuit Development at Brookhaven National Laboratory. (57-398)....365–8
Biringer, Paul P. Disc....293
Biringer, P. P.; G. W. Dick, G. R. Slemon. Two-Phase Motor Drive Using Static Frequency-Doubler Amplifiers. (57-631)....497–503; disc. 504
Black, K. G.; R. J. Noorda. Analog Computer Study of Wind Tunnel Drive. (57-1008)....745–50; disc. 750
Bloecker, W. E.; G. A. Caldwell. The Work of the International Telecommunication Union in the Field of Telephony. (57-785)....543–52
Boffi, Luiz V.; V. B. Haas, Jr. Analog Computer Representation of Alternators for Parallel Operations. (57-150)....153–7; disc. 157
Bonneville, S. The Canadian Transcontinental Microwave System. (57-706)....473–7
Bourne, H. C.; T. T. Kadota. Operational Magnetic Amplifiers with Multiple Control Windings. (57-772)....515–20
Bourne, H. C., Jr.; D. Nitzan. Dynamic Operation of Magnetic Amplifiers with Real-Core and Rectifier Functions. (57-94)....646–55
Bowers, F. K. What Use Is Delta Modulation to the Transmission Engineer? (57-136)....142–7
Brant, C. M. A Canadian Point of View on Radio-Frequency Spectrum Management. (57-647)....455–61
Bruce, F. M. Disc....313
Brustman, J. A.; J. Agin. The BIZMAC II Tapefile. (57-674)....728–31

C

Caldwell, G. A.; W. E. Bloecker. The Work of the International Telecommunication Union in the Field of Telephony. (57-785)....543–52
Camp, H. R.; H. W. Henkels, J. R. Fortier, W. J. Shilling, R. E. Smith. Silicon Rectifiers in Brushless Machinery. (57-793)....620–6
Chang, S. S. L. Statistical Design Theory for Strictly Digital Sampled Data Systems. (57-1061)....702–08; disc. 709
Chastain, J. W.; C. V. Weaver, C. K. Smith. Development of a Thermal-Neutron Flux-Measuring Instrument. (57-406)....573–7
Cherry, E. Colin. Disc....496
Chilberg, G. L. Buried Cable Telephone Systems. (57-140)....130–5
Chin, T. H.; T. J. Higgins. Equations for the Inductances and Current Distribution of Multiple-Conductor Single-Phase and Polyphase Busses. (57-645)....553–8
Choksy, N. H. Disc....533
Cioffi, P. P. Rectilinearity of Electron-Beam Focusing Fields from Transverse Component Determinations. (56-518)....15–19
Clabaugh, R. G.; E. T. Heald. A Predicted Wave-Signaling Phase-Shift Telegraph System. (57-65)....316–19; disc. 319
Clabaugh, R. G.; R. R. Mosier. Kineplex, A Bandwidth Efficient Binary Transmission System. (57-1042)....723–7; disc. 728
Clark, E. W.; H. E. Albright. Improved Direct-Acting Electric Recording Instruments. (57-220)....215–21; disc. 221
Clark, J. H. A New Video Differential Phase-and-Gain Equalizer. (57-919)....674–6
Clemency, W. F.; F. F. Romanow, A. F. Rose. The Bell System Speakerphone. (57-139)....148–53
Cline, Jack F. Small-Signal Heterodyne Mixers with Excessive Injection Amplitudes. (57-1027)....739–45
Coker, Lawrence F. The Development of a New Method of Circuit Analysis in Ladder Networks. (57-104)....158–60
Combellick, T. A.; M. E. Ferguson. Noise Considerations on Toll Telephone Microwave Radio Systems (57-59)....67–70

Corey, H. E., Jr. Disc.......669

Critchlow, D. L.; L. A. Finzi. Dynamic Core Behavior and Magnetic Amplifier Performance. (56-272).......229–40

D

Dahlin, Bjorn E. Radiation-Flux Conditions in Radioactive Media with Applications to Radiation Monitoring. (57-401).......391–6

Davis, Dean W. Characteristics and Applications of the Iatron Storage Tube. (57-7).......47–53

DeFalco, Francis R.; A. E. Klickman. A Wire-Activation Technique for Reactor-Flux-Profile Measurements. (57-405).......428–30

Denholm, A. S. Disc.......313

DeSautels, Albert N. Transistorized Phase Discriminators. (57-31).......19–26

Diamond, B. S. Multiplex Telegraph Equipment for Radio and Submarine Cable Circuits. (57-64).......175–82; disc. 182

Dick, G. W.; P. P. Biringer, G. R. Slemon. Two-Phase Motor Drive Using Static Frequency-Doubler Amplifiers. (57-631).......497–503; disc. 504

Diebold, E. J. Temperature Rise of Solid Junctions Under Pulse Load. (57-744).......593–8

Dinnin, A. J. Light-Route Microwave Systems in Canada. (57-707).......488–92

Distler, R. J.; C. Sturzenbecker. A Magnetic Amplifier Voltage-Current Overload Protection Circuit. (57-771).......602–06

Doak, F. C.; J. R. Tennet. Development in Communications in the Lower St. Lawrence Valley. (57-10).......26–39

Dortort, I. K. A New Voltage-Divider Circuit for Semiconductor Rectifiers. (57-395).......356–8; disc. 358

Dunkin, J. W.; G. R. Partridge, R. L. Anderson, R. C. Geldmacher. A 14-Channel Displacement Measuring Device Utilizing Magnetic and Paper Tape Recording. (57-662).......461–7

E

Early, H. C.; R. C. Walker. Economics of Multimillion-Joule Inductive Energy Storage. (57-79).......320–4; disc. 324

Edwards, F. S. Disc.......314

Edwards, G. D. Disc.......669

Egan, T. F.; H. W. Hermance. The Examination of Electric Contacts by the Plastic Replica Method. (57-1095).......756–62

Emling, J. W. General Aspects of Hands-Free Telephony. (57-137).......201–05

Estoppey, R. E. Disc.......327

Evers, H. G.; J. R. Wylie. The Automatic Line Concentrator and Subscriber Telephone Applications. (57-47).......6–9

F

Ferguson, M. E.; T. A. Combellick. Noise Considerations on Toll Telephone Microwave Radio Systems. (57-59).......67–70

Fich, S.; J. J. Slade, Jr., L. F. Nanni, D. A. Molony. Moment Detection and Coding. (56-963).......275–9

Fiegel, H. E.; W. A. Keen, Jr. Factors Influencing the Sparkover Voltage of Asymmetrically Connected Sphere Gaps. (57-20).......307–12; disc. 315

Finzi, L. A.; D. L. Critchlow. Dynamic Core Behavior and Magnetic Amplifier Performance. (56-272).......229–40

Fisher, F. A.; A. F. Rohlfs, J. S. Kresge. The Response of Resistance Voltage Dividers to Steep-Front Impulse Waves. (57-353).......634–42; disc. 645

Foley, A. H. A Self-Balancing Transformer-Core-Loss Bridge. (57-559).......567–73

Fortier, J. R.; H. W. Henkels, H. R. Camp, A. J. Shilling, R. E. Smith. Silicon Rectifiers in Brushless Machinery. (57-793).......620–6

Foust, C. M.; B. V. Bhimani. A New Voltage Time Recorder for Transient Measurements. (57-214).......253–61; disc. 262

Freeman, Herbert. Cycle- and Delay-Time Considerations in a Real-Time Digital Computer. (57-677).......588–93

Frey, H. A. Disc.......314

G

Gano, J. J. The Elimination of Motor-Generator Sets as Transient Buffers for SAGE Computers. (57-676).......468–73

Garbarino, H. L. Magnetic Circuits with D-C Magnetization and High Alternating Induction. (57-599).......338–48

Gaudet, G.; J. B. Pringle. A Statistical Approach to Telephone Transmission Maintenance. (57-784).......663–9

Geldmacher, R. C.; G. R. Partridge, J. W. Dunkin, R. L. Anderson. A 14-Channel Displacement Measuring Device Utilizing Magnetic and Paper Tape Recording. (57-662).......461–7

Geyger, William A. Self-Balancing Magnetic Amplifiers of the Differential-Feedback Type. (57-92).......39–46

Goldman, O. H.; J. E. Ross. Mechanical Features of 561 Subscriber Carrier. (57-1017).......677–81

Gray, T. S.; W. M. Grim, Jr., F. S. Replogle, Jr., R. H. Spencer. Solid-State Neutron-Flux Measuring System. (57-407).......368–72

Gray, T. S.; A. B. VanRennes, J. C. Simons, Jr. Dynamics and Control of Nuclear Power Plants. (57-390).......279–85

Grim, W. M., Jr.; T. S. Gray, F. S. Replogle, Jr., R. H. Spencer. Solid-State Neutron-Flux Measuring System. (57-407).......368–72

Griscom, S. B.; B. L. Lloyd, A. R. Hileman. Disc.......642

H

Haas, Vinton B., Jr.; L. V. Boffi. Analog Computer Representation of Alternators for Parallel Operations. (57-150).......153–7; disc. 157

Halacsy, Andrew A. Disc.......503

Halloran, James J. Deficiency of Standards for High-Voltage High-Power Rectifier Transformers. (57-1014).......690–2; disc. 694

Hannula, Fred W. The BTU Computer: An Advanced Data System for Industry. (57-222).......222–5

Har'el, Abraham. Transistor 2-Terminal Switches. (57-106).......328–38

Harrington, H. J.; J. Zaborszky. A Describing Function for Multiple Nonlinearities Present in Electrohydraulic Control Valves. (57-107).......183–90

Harrington, H. J.; J. Zaborszky. Generalized Charts of the Effects of Nonlinearities in Electrohydraulic Control Valves. (57-108).......191–8

Hayes, J. E. The Development of Television in Canada. (57-665).......482–4

Heald, E. T.; R. G. Clabaugh. A Predicted Wave-Signaling Phase-Shift Telegraph System. (57-65).......316–19; disc. 319

Hellerman, Herbert. A Generalized Theory of Transistor Bias Circuits. (57-1023).......694–7

Henkels, H. W.; H. R. Camp, J. R. Fortier, W. J. Shilling, R. E. Smith. Silicon Rectifiers in Brushless Machinery. (57-793).......620–6

Hermach, Francis L. Disc.......249

Hermance, H. W.; T. F. Egan. The Examination of Electric Contacts by the Plastic Replica Method. (57-1095).......756–62

Hewitt, G. W. The Charging of Small Particles for Electrostatic Precipitation. (57-90).......300–06; disc. 306

Hicks, W. Wesley; J. C. Beckett. The Control of Air Ionization and Its Biological Effects. (57-73).......108–11

Higgins, Thomas J.; T. H. Chin. Equations for the Inductances and Current Distribution of Multiple-Conductor Single-Phase and Polyphase Busses. (57-645).......553–8

Hileman, A. R.; S. B. Griscom, B. L. Lloyd. Disc.......642

Hill, G. Leslie. Disc.......692

Hingston, A. J.; W. F. Santelmann, Jr. Design of Long Passive Oscilloscope Probes to Simplify Maintenance of Large Digital Computers. (57-542).......359–62

Hollitscher, H. Disc.......503

Honnell, Pierre M.; D. F. Wann. Disc.......160

Housman, B.; M. M. Astrahan, J. F. Jacobs, R. P. Mayer, W. H. Thomas. The Logical Design of a Digital Computer for a Large Scale Real-Time Application. (57-200).......71–5

Hu, K. C.; Y. H. Ku. Circuit for a Magnetic Subharmonic Pulser. (57-99).......137–41

Hubbard, Robert M. Magnetic Amplifier Analysis and Application Using Block-Diagram Techniques. (57-766).......578–88

Hunt, T. W. Electrical Problems in Electrical Burn-Out Testing of Nuclear Fuel Elements. (57-404).......400–06

Huntley, H. R. Where We Are and Where We Are Going in Telephone Transmission. (56-780).......54–63

Hylten-Cavallius, N. Disc.......312, 644

J

Jacob, G. R. Dynamics of Electronic Self-Balancing Systems. (57-17).......166–74

Jacobs, I. M. Safety Systems for Nuclear Power Reactors. (57-906).......670–3

Jacobs, J. F.; M. M. Astrahan, B. Housman, R. P. Mayer, W. H. Thomas. The Logical Design of a Digital Computer for a Large Scale Real-Time Application. (57-200).......71–5

Jamieson, D. F.; B. R. Stachiewicz, R. L. Layburn. A New 5-Channel Carrier System for Subscriber Use Employing Transistors. (56-802).......385–90

Johannessen, Paul R.; R. B. Mark, W. X. Johnson. Magneticalley Keyed, Phase-Sensitive Demodulators. (56-757).......1–6

Johnson, Wheeler X.; R. B. Mark, P. R. Johannessen. Magnetically Keyed, Phase-Sensitive Demodulators. (56-757).......1–6

Jones, A. R.; R. L. Witzke. Some Operational Problems of the Nuclear Power Plant. (57-396).......373–7

K

Kadota, T. T.; H. C. Bourne. Operational Magnetic Amplifiers with Multiple Control Windings. (57-772).......515–20

Karraker, I. O. Quality Assurance—An Essential Part of the Western Electric Manufacturing Process. (57-808).......540–2

Keen, W. A., Jr.; H. E. Fiegel. Factors Influencing the Sparkover Voltage of Asymmetrically Connected Sphere Gaps. (57-20).......307–12; disc. 315

Kellogg, H. L. Disc.......508

Kernick, Andress; H. A. Perkins, Jr. Magnetic Output Amplifiers for Digital Control. (57-774).......504–08; disc. 508

Keyser, G. M.; J. R. Leslie. Advanced Data System in the Power Industry. (57-221).......206–10

Kikuchi, Tadashi. Analysis of the Full-Wave Magnetic Amplifier Circuits Considering the Change of the Width of the Dynamic Hysteresis Loop. (57-97).......241–9

Kikuchi, T.; K. Murakami. On the Input Power of the Half-Wave Magnetic Amplifier Circuit. (57-5).......10–14

Klickman, Alton E.; F. R. DeFalco. A Wire-Activation Technique for Reactor-Flux-Profile Measurements. (57-405).......428–30

Krakauer, Stewart M. Elimination of the Ballast Resistor in D-C Power Supplies. (57-78).......101–08

Krall, Albert D. The Operation of the Current-Type Self-Balancing Magnetic Amplifier. (57-765).......380–4

Kresge, J. S.; J. G. Anderson. An Electronic Electrometer as a Versatile Corona Detector. (57-632).......449–54

Kresge, J. S.; A. F. Rohlfs, F. A. Fisher. The Response of Resistance Voltage Dividers to Steep-Front Impulse Waves. (57-353).......634–42; disc. 645

Krishna, G.; P. V. Rao. The Transformer Analog Computer. (57-1063).......732–8

Ku, Y. K.; K. C. Hu. Circuit for a Magnetic Subharmonic Pulser. (57-99).......137–41

L

Landry, J. G.; J. F. Scamman. An Improved Single-Phase Watt-Hour Meter. (57-651).......438–9

Layburn, R. L. Engineering Features and Field Trial Performance of a New Subscriber Carrier System. (57-1018).......681–7

Layburn, R. L.; B. R. Stachiewicz, D. F. Jamieson. A New 5-Channel Carrier System for Subscriber Use Employing Transistors. (56-802).......385–90

Lee, Thomas, H. Forces and Stresses in an Electromagnetic Field. (57-109).......267–71; disc. 274

Lehtonen, E. W. Disc.......86

Leonhard, W. Disc.......533

Leonhard, W. Speed Control of a D-C Motor Using a Magnetic Amplifier. (57-93).......112–19

Leslie, J. R.; G. M. Keyser. Advanced Data System in the Power Industry. (57-221).......206–10

Lichtenstein, Roland M. Disc.......271

Linck, H. Disc.......261

Lloyd, B. L.; S. B. Griscom, A. R. Hileman. Disc.......642

Löf, John L. C. Solution of Simultaneous Equations by Block Elimination. (57-741).......509–14

Lones, W.; J. E. Binns, D. G. Pitcher, M. Melice. Safety Circuit Development at Brookhaven National Laboratory. (57-398).......365–8

Lynch, Robert D.; G. W. Penney. Measurements of Charge Imparted to Fine Particles by a Corona Discharge. (57-89).......294–9; disc. 306

M

Macintyre, J. R.; R. G. Ballard. Mass Flowmeter with Retarding Disk Integration. (57-113).......127–30

Manning, M. L. New Concepts in Evaluating Electric Equipment Insulation Systems—The Problems Encountered. (57-392).......348–55

Manteuffel, E. W. Disc.......503

Manteuffel, E. W.; R. O. McCary. The D-C Controlled A-C Voltage Source, A New Magnetic Amplifier. (57-768).......562–6

Mark, Robert B.; W. X. Johnson, P. R. Johannessen. Magnetically Keyed, Phase-Sensitive Demodulators. (56-757).......1–6

Maunder, S. T. Disc.......692

Mayer, R. P.; M. M. Astrahan, B. Housman, J. F. Jacobs, W. H. Thomas. The Logical Design of a Digital Computer for a Large Scale Real-Time Application. (57-200).......71–5

McCary, R. O.; E. W. Manteuffel. The D-C Controlled A-C Voltage Source, A New Magnetic Amplifier. (57-768).......562–6

McClatchie, F. F. A Modified Microwave System for Color Television Transmission. (57-918)...626–33
McDavitt, M. B. 6,000-Megacycle-Per-Second Radio Relay System for Broad-Band Long-Haul Service in the Bell System. (57-1044)..............715–22
McMurray, William. Magnetic Frequency Multipliers and Their Rating:
  Part II. Frequency Sextuplers. (57-91)..289–93; disc. 293
Melice, M.; J. E. Binns, W. Lones, D. G. Pitcher. Safety Circuit Development at Brookhaven National Laboratory. (57-398)..............365–8
Mertz, Pierre. Information Theory Impact on Modern Communications. (57-635)................431–7
Miller, Charles J., Jr. Disc...........................313
Miller, J. H. An Inductronic Double Bridge. (57-111)..................................198–200
Miller, W. A. A Narrow-Band Experimental FM Mobiletelephone System. (57-58)........98–100
Mitchell, K. M. Disc...........................221
Molony, D. A.; J. J. Slade, Jr., L. F. Nanni, S. Fich. Moment Detection and Coding. (56-963)...275–9
Montgomery, G. Franklin; C. Stansbury. The Clamp-Type A-C Microammeter. (57-114)..................................249–53; disc. 253
Moore, J. B. Disc...........................319
Moore, John B. Disc...........................182
Mosier, R. R.; R. G. Clabaugh. Kineplex, A Bandwidth Efficient Binary Transmission System. (57-1042)......................723–7; disc. 728
Mostafa, Abd El-Samie. Simultaneous Pulled Oscillations in a Triode Oscillator Incorporating Two Oscillatory Circuits. (57-77)..............120–7
Murakami, K.; T. Kukuchi. On the Input Power of the Half-Wave Magnetic Amplifier Circuit. (57-5)..10–14

N

Nanni, L. F.; J. J. Slade, Jr., S. Fich, D. A. Molony. Moment Detection and Coding. (56-963)..275–9
Nesline, Frederick W., Jr. Disc..................708
Newell, E. L. Loading Coils for Ocean Cables. (57-705)................................478–82
Niehaus, Udo K.; R. E. Wall, Jr. Russian to English Machine Translation with Simple Logical Processing. (57-1062).........................709–14
Nitzan, D.; H. C. Bourne, Jr. Dynamic Operation of Magnetic Amplifiers with Real-Core and Rectifier Functions. (56-94).......................646–55
Noorda, R. J.; K. G. Black. Analog Computer Study of Wind Tunnel Drive. (57-1008)..745–50; disc. 750

O

Olding, N. R. The Canadian Broadcasting Corporation's Radio Service in Canada. (57-659)..558–61

P

Park, J. H. Disc.........................261,645
Parry, Charles A. Design Factors for the Optimization of Multichannel Radio Systems. (57-60)...606–20
Partridge, G. R.; J. W. Dunkin, R. L. Anderson, R. C. Geldmacher. A 14-Channel Displacement Measuring Device Utilizing Magnetic and Paper Tape Recording. (57-662).....................461–7
Penney, Gaylord W.; R. D. Lynch. Measurements of Charge Imparted to Fine Particles by a Corona Discharge. (57-89).............294–9; disc. 306
Perkins, H. A., Jr.; A. Kernick. Magnetic Output Amplifiers for Digital Control. (57-774)..................................504–08; disc. 508
Pitcher, D. G.; J. E. Binns, W. Lones, M. Melice. Safety Circuit Development at Brookhaven National Laboratory. (57-398)..............365–8
Pitman, J. B. Disc...........................358
Plotkin, Sheldon. Discontinuous Transition Time Between Stable States in Ferroresonant Circuits. (56-789).................................410–21
Powers, K. H.; H. Staras. Some Relations Between Television Picture Redundancy and Bandwidth Requirements. (57-722). [illegible]92–6; disc. 496
Pringle, J. B.; G. Gaudet. [illegible] Approach to Telephone Transmission Maintenance. (57-784)...663–9

R

Raber, Robert A. Magnetically Regulated Reference Power Supply. (57-770)..................534–7
Rao, P. Venkata; G. Krishna. The Transformer Analog Computer. (57-1063)....................732–8
Replogle, F. S., Jr.; T. S. Gray, W. M. Grim, Jr., R. H. Spencer. Solid-State Neutron-Flux Measuring System. (57-407)........................368–72
Replogle, Frank S., Jr.; D. R. Whitehouse. Acoustic Ionization Detector. (57-408)............406–09
Rodek, V. The Hands-Free Telephone. (57-138)...210–14
Rohlfs, A. F.; J. S. Kresge, F. A. Fisher. The Response of Resistance Voltage Dividers to Steep-Front Impulse Waves. (57-353)..........634–42; disc. 645
Romanow, F. F.; W. F. Clemency, A. F. Rose. The Bell System Speakerphone. (57-139)..........148–53
Rose, A. F.; W. F. Clemency, F. F. Romanow. The Bell System Speakerphone. (57-139)......148–53
Ross, Hugh C. Disc...........................693
Ross, J. E.; O. H. Goldman. Mechanical Features of 561 Subscriber Carrier. (57-1017)........677–81
Rowe, William D.; G. H. Royer. Transistor NOR Circuit Design. (57-196)....................263–7
Royer, Goerge H.; W. D. Rowe. Transistor NOR Circuit Design. (57-196)....................263–7
Rubinoff, Morris. A New Family of Transistor Switching Circuits. (57-105)..................286–9

S

Santelmann, W. F., Jr.; A. J. Hingston. Design of Long Passive Oscilloscope Probes to Simplify Maintenance of Large Digital Computers. (57-542)...359–62
Sauer, L. E. Disc...........................324
Scamman, J. F.; J. G. Landry. An Improved Single-Phase Watt-Hour Meter. (57-651)........438–9
Scheinman, A. H. A Numerical-Graphical Method for Synthesizing Switching Circuits. (57-1019)..687–9
Seegmiller, W. R. High-Frequency Magnetic Servo Amplifier for Air-Borne Control Systems. (57-775)..526–9
Shilling, W. J.; H. W. Henkels, H. R. Camp, J. R. Fortier, R. E. Smith. Silicon Rectifiers in Brushless Machinery. (57-793)....................620–6
Sias, F. R. A Resistance-Temperature Detector for Nuclear Reactor Service. (57-901).........363–5
Simons, J. C., Jr.; A. B. VanRennes, T. S. Gray. Dynamics and Control of Nuclear Power Plants. (57-390)..................................279–85
Slade, J. J., Jr.; L. F. Nanni, S. Fich, D. A. Molony. Moment Detection and Coding. (56-963)...275–9
Slemon, G. R.; G. W. Dick, P. P. Biringer. Two-Phase Motor Drive Using Static Frequency-Doubler Amplifiers. (57-631)..........497–503; disc. 504
Smith, C. K.; C. V. Weaver, J. W. Chastain. Development of a Thermal-Neutron Flux-Measuring Instrument. (57-406)........................573–7
Smith, D. M. Design of an Improved 2-Stator Watt-Hour Meter. (57-792)..................537–40
Smith, R. E.; H. W. Henkels, H. R. Camp, J. R. Fortier, W. J. Shilling. Silicon Rectifiers in Brushless Machinery. (57-793)....................620–6
Smith, Richard, E.; E. M. Williams. Phenomena Accompanying Transient Low-Voltage Discharges in Liquid Dielectrics:
  Part II. (57-2)................................93–7
Spencer, R. H.; T. S. Gray, W. M. Grim, Jr., F. S. Replogle, Jr. Solid-State Neutron-Flux Measuring System. (57-407).......................368–72
Stachiewicz, B. R.; R. L. Layburn, D. F. Jamieson. A New 5-Channel Carrier System for Subscriber Use Employing Transistors. (56-802).........385–90
Stansbury, Carroll; G. F. Montgomery. The Clamp-Type A-C Microammeter. (57-114)..................................249–53; disc. 253
Staras, H.; K. H. Powers. Some Relations Between Television Picture Redundancy and Bandwidth Requirements. (57-722).........492–6; disc. 496
Stone, R. S. Recent Developments in Nuclear Instrumentation at the Knolls Atomic Power Laboratory. (57-399)................................396–9
Strong, E. M.; E. J. Warchol, S. E. Benson. Microwave Radio Communication Under Severe Winter Conditions on the Bonneville Power Administration System. (57-915)......................655–62
Sturzenbecker, C.; R. J. Distler. A Magnetic Amplifier Voltage-Current Overload Protection Circuit. (57-771)................................602–06
Sunde, E. D. Disc...........................727
Sze, Wilbur C. Measurement of Voltage Ratio at Audio Frequencies. (57-648)..............444–9

T

Tennet, J. R.; F. C. Doak. Development in Communications in the Lower St. Lawrence Valley. (57-10)..26–39
Thomas, W. H.; M. M. Astrahan, B. Housman, J. F. Jacobs, R. P. Mayer. The Logical Design of a Digital Computer for a Large-Scale Real-Time Application. (57-200)....................71–5

U

Urkowitz, Harry. Signal-to-Noise Ratios in Strong Carrier FM Systems. (57-61)..............599–602
Usher, Theron, Jr. A New Application of the Hurwitz-Routh Stability Criteria. (57-742).530–3; disc. 533
Usher, Theron, Jr. A Nonlinear Low-Frequency Instability Phenomenon in Audio Amplifiers. (57-1029)..698–701
Usher, Theron, Jr. Design Relationships for Iron-Core Filter Chokes. (57-743)..................484–7

V

Vallese, L. M. Transient Analysis of Second-Order Flip-Flops. (57-81).....................161–6
Vallese, L. M. Upper Limits of Output Power in Vacuum-Tube and Transistor A-C Amplifiers. (57-74)................................87–92
Van Rennes, A. B.; J. C. Simons, Jr., T. S. Gray. Dynamics and Control of Nuclear Power Plants. (57-390)................................279–85
Vrana, Norman M. Compensation for Instrument Losses in Circuits Containing a Wattmeter, Voltmeter, and Ammeter. (57-43)...325–7; disc. 327

W

Walker, R. C.; H. C. Early. Economics of Multimillion-Joule Inductive Energy Storage. (57-79)..................................320–4; disc. 324
Wall, Robert E., Jr.; U. K. Niehaus. Russian to English Machine Translation with Simple Logical Processing. (57-1062).........................709–14
Wann, Donald F.; P. M. Honnell. Disc...........160
Warchol, E. J.; E. M. Strong, S. E. Benson. Microwave Radio Communication Under Severe Winter Conditions on the Bonneville Power Administration System. (57-915)......................655–62
Warman, W. C. Closed-Circuit Network for Educational Television. (57-668)..............440–4
Weaver, C. V.; C. K. Smith, J. W. Chastain. Development of a Thermal-Neutron Flux-Measuring Instrument. (57-406)........................573–7
Weeks, John G. Dynamic Gauging and Automatic Adjusting of Wire Spring Relays. (57-810)..520–6
Westendorp, W. F. Bias Systems for Resonance Transformer Million-Volt Electron Beam Generators. (57-1057)...............................751–5
Whitehouse, David R.; F. S. Replogle, Jr. Acoustic Ionization Detector. (57-408)............406–09
Williams, Everard M.; R. E. Smith. Phenomena Accompanying Transient Low-Voltage Discharges in Liquid Dielectrics. Part II. (57-2).......93–7
Willis, F. H. Some Results with Frequency Diversity in a Microwave Radio System. (56-1001)......63–7
Witzke, R. L.; A. R. Jones. Some Operational Problems of the Nuclear Power Plant. (57-396)..373–7
Wouk, Victor. Disc...........................693
Wylie, J. R.; H. G. Evers. The Automatic Line Concentrator and Subscriber Telephone Applications. (57-47)..................................6–9

Z

Zaborszky, J.; H. J. Harrington. A Describing Function for Multiple Nonlinearities Present in Electrohydraulic Control Valves. (57-107).......183–90
Zaborszky, J.; H. J. Harrington. Generalized Charts of the Effects of Nonlinearities in Electrohydraulic Control Valves. (57-108).................191–6

Magnetically Keyed, Phase-Sensitive Demodulators

ROBERT B. MARK
NONMEMBER AIEE

WHEELER X. JOHNSON
ASSOCIATE MEMBER AIEE

PAUL R. JOHANNESSEN
ASSOCIATE MEMBER AIEE

THE purpose of this paper is to present several keyed, phase-sensitive demodulator circuits in which saturable reactors, transistors, and silicon diodes are used in place of vacuum tubes to obtain increased reliability and ruggedness, small physical size (see Fig. 1), and low noise level.

The theory of keyed, phase-sensitive demodulators is described in detail in reference 1. Only a brief description of the basic operation will be given here. Demodulation is effected by sampling the a-c input voltage once per cycle. This sampling occurs during a short interval of time at the peak of the a-c wave form. For a 400-cycle-per-second system typical values of the sampling interval range from 1 to 50 μsec (microseconds). Between samplings, the sampled peak value of the a-c voltage wave is stored on a capacitor. The sampling process is performed by an electronic switch composed of diodes or transistors and controlled by a voltage pulse. The switch is normally open and closes only during the pulse interval. In this paper the electronic switch is called the "gating circuit" and the circuit which generates the pulses controlling the switch is called the "pulse-forming circuit." Fig. 2 shows the gating circuit schematically. Among currently available methods of demodulation, this sampling method provides the smallest ripple for a given speed of response.

The three distinct subtypes of demodulators to be discussed are the diode demodulator, which is characterized by inherent linearity; the magnetic amplifier demodulator, which can operate from sources of relatively high impedance; and the transistor demodulator, which combines good linearity with low noise level.

Although details of operation differ from one subtype to the next, they are similar in over-all operation and each demodulator comprises two parts, a pulse-forming circuit and a gating circuit. The same pulse-forming circuit, with minor modifications, is common to all the demodulators to be described, and is treated separately in the following section. The ensuing sections describe the three different types of gating circuits previously mentioned.

The Pulse-Forming Circuit

The pulse-forming circuit[2] is shown in Fig. 3 and operates as follows. For the positive half-cycles of the line voltage e_s, reactor I absorbs all the available voltage until it saturates, because it has many more turns and a much smaller magnetizing current than reactor II. Saturation is very sharp because the reactor core material is of the rectangular hysteresis-loop type (Orthonol or Deltamax).

The sum of the line voltage e_s and the voltage across capacitor C_1 now appears across reactor II, which saturates very quickly (in 10 to 25 μsec for the circuits treated here), because of its relatively few turns.

After reactor II saturates, the capacitor C_1 is connected directly across the line voltage. If reactor I saturates at the peak of the line voltage, capacitor C_1 will charge to the peak of the line voltage after saturation of reactor II. The voltage across reactor I is the difference between the capacitor voltage and the line voltage.

The voltage e_{RI} across reactor I (Fig. 4) reverses polarity if the discharge time-constant R_1C_1 is sufficiently large. When R_1 is increased, the voltage across capacitor C_1 approaches a square wave and the volt-time area applied to reactor I (the shaded area in Fig. 4) tends to increase. However, the volt-time area required to saturate reactor I is fixed by the number of turns. An increase in the R_1C_1 time-constant therefore will cause the core to saturate earlier in the cycle.

This phenomenon is utilized to adjust the firing of reactor I so that it occurs exactly at the 90-degree point of the reference wave. During the negative half-cycles of the reference voltage, the same chain of events occurs except that the height of the negative pulses is reduced by conduction of the *1N92* rectifier. For reasons to be developed later, reduction of the negative pulses increases the maximum output of the demodulators. However, if this reduction is carried too far by excessively reducing the value of resistor R_2, reactor II will not reach negative saturation or full reset, and the height of the positive pulse as well as the negative pulse will be reduced. The practical minimum for the negative pulse height is about one fifth the height of the positive pulse.

The pulse-forming circuit output wave form is shown in Fig. 5, with details of the positive and negative pulses on an expanded scale. Fig. 6 shows the positive pulse structure, and has been labeled for the following discussion. The rise time of this pulse is caused by the slight rounding of the hysteresis loop and by the saturated inductance of reactor I. For the circuit shown in Fig. 3, the rise time is approximately 8 μsec. This rise time may be increased, if desired, by adding a capacitor across the input winding of reactor II.

In connection with the operation of the circuits to be described, it is helpful to note the effect of decreasing the value of resistor R_2. A reduction of R_2 will decrease the amount of reset of reactor II, which in turn will reduce the width of the pulse. As seen from Fig. 6, the pulse width can be reduced until point *A* is reached without materially changing the height of the pulse. Further reduction in the pulse width, however, will result in a rapid change in the height. The rate of reduction is governed by the rise time of the pulse.

At this point it is convenient to investigate some of the undesirable components appearing in the output voltage of the pulse-forming circuit. These

---

Paper 56-757, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956, and re-presented at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted November 9, 1955; made available for printing December 7, 1956.

ROBERT B. MARK, WHEELER X. JOHNSON, and PAUL R. JOHANNESSEN are with the Massachusetts Institute of Technology, Cambridge, Mass.

The research reported in this document was supported jointly by the Department of the Army, the Department of the Navy, and the Department of the Air Force under Air Force Contract No. *AF19(122)-458* with Lincoln Laboratory, Massachusetts Institute of Technology, and executed under subcontract by the Massachusetts Institute of Technology Servomechanisms Laboratory, Department of Electrical Engineering. It is published for technical information only and does not represent recommendations or conclusions of the sponsoring agencies.

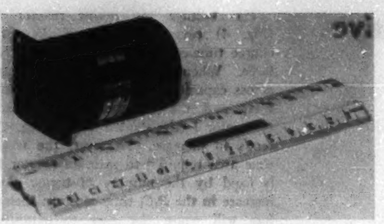


Fig. 1 (left). Packaged magnetic demodulator

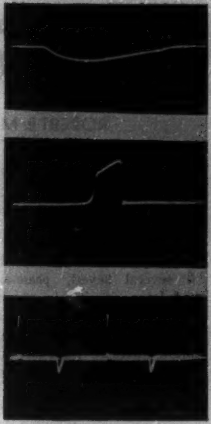


Fig. 5 (right). Pulse-forming circuit output wave form

henceforth will be referred to as "noise." A small amount of actual noise appears as a result of random variations of the diode and core properties from cycle to cycle, but in general the more important sources of noise are the line voltage and the negative pulses. A small fraction of the line voltage appears in the output of the pulse-forming circuit, because the impedance of reactor II is never entirely negligible. A larger component of the line voltage is transferred to the output by coupling or pickup.

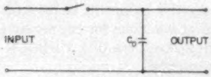
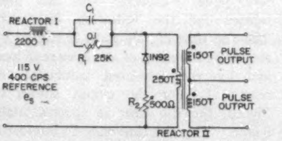


Fig. 2. Gating circuit-schematic representation



REACTOR I

CORE: MAGNETICS, INC. TOROIDAL CORE NO. 50176-2A
WINDING: 2200T, NO. 36 AWG ENAMELLED WIRE

REACTOR II

CORE: MAGNETICS, INC. TOROIDAL CORE NO. 50153-1A
PRIMARY WINDING: 250T, NO. 36 AWG ENAMELLED WIRE
SECONDARY WINDING: 300T, CENTER TAPPED, NO. 38 AWG ENAMELLED WIRE

Fig. 3. Pulse-forming circuit

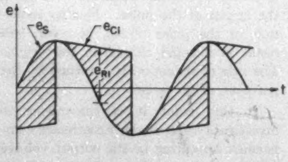


Fig. 4. Voltage wave form across capacitor C_1

The Diode Demodulator

The first of the demodulators to be described is the diode demodulator, for which the circuit diagram is shown in Fig. 7. For clarity, the pulse-forming circuit is omitted. It will also be omitted in most of the following diagrams of demodulators.

The operation of this circuit can be explained as follows. With zero input voltage, the arrival of a positive pulse results in charging the bias capacitors C_1 and C_2 to a value of voltage equal to the pulse height, and thereby prevents the diodes from conducting except during the pulse period. In the period between one positive pulse and the next, these capacitors gradually discharge through resistors R_1 and R_2 to about 80 per cent of the pulse height. This discharge of the bias capacitors between pulses is necessary because it permits unblocking of the diodes during the succeeding pulse interval. However, if the discharge (or voltage decay) is unequal for the two capacitors, an output d-c component will appear for zero input voltage.

A signal applied to the input of the demodulator will meet a high impedance path to the output capacitor C_0 except during the pulses, which occur at the 90-degree phase point or peak value of the reference voltage. During the pulses, the diodes are unblocked by the pulse voltages, and the input voltage is presented with a low-impedance path to the output capacitor. If the source resistance and the forward diode resistance are sufficiently small, the time constant for charging the output capacitor can be made much smaller than the pulse duration. This allows the output voltage to reach substantially the full value of the peak input voltage in the duration of one pulse. This action is here defined as "1-cycle response." If, however, the charging resistance or the value of output capacitance is too large, the response to a step of input voltage will resemble a series of small steps superimposed upon an ordinary exponential rise. This is demonstrated in Fig. 8, which pictures the step response for different values of the output capacitance. Response to a sinusoidal input is shown in Fig. 9. Because of the timing of the pulses, the output voltage will be equal to

$$E_0 = E_i \cos \phi$$

where ϕ is the phase angle of the input signal with respect to the reference

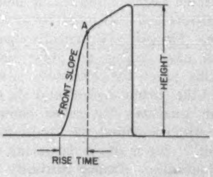


Fig. 6. Positive pulse structure

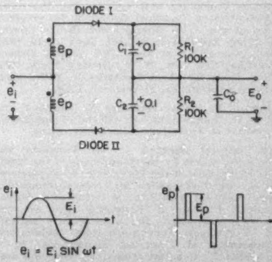


Fig. 7. Circuit-diode demodulator

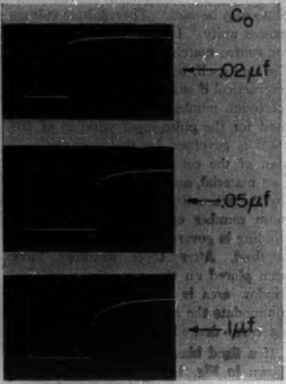


Fig. 8 (left). Response to step-voltage input

Fig. 9 (above). Response to sinusoidal input

voltage. Consequently, the demodulator will not respond to signals in quadrature with the reference voltage.

To obtain 1-cycle response, the charging time constant should be less than one fifth of the pulse width. For instance, for a pulse width of 25 μsec the charging time constant must be 5 μsec or less. If the output capacitor C_0 is 0.01 μf (microfarad) the maximum value of source impedance R_s in series with the forward impedance of the diode R_f is

$$(R_f+R_s)_m=\frac{5\times10^{-6}}{0.01\times10^{-6}}=500 \text{ ohms}$$

Because R_f is generally much smaller than 500 ohms, it may be neglected and 500 ohms may be taken as the limit of permissible source resistance for 1-cycle response. However, this low-impedance value need only be maintained during the pulse, permitting thereby the use of a circuit such as is shown in Fig. 10 to provide low effective source impedance. Phase shift of the input signal is not harmful, because the quadrature-rejection properties of the demodulator convert it to a mere loss of gain that is more than compensated by the gain of the tube. If the cathode follower connection were used to reduce the source impedance of the tube, the gain would be lost.

Because the output noise is inversely proportional to output capacitance, while the charging time-constant is directly proportional to it, the value chosen for output capacitance implies a compromise between noise and speed of response.

The maximum input voltage of the diode demodulator is limited either by the possibility of backward diode current if the Zener voltage is exceeded, or by forward diode current between pulses if the voltage across one of the diodes becomes positive. By reference to Fig. 7, the maximum input voltage of the diode demodulator may be computed.

In the Appendix this computation is performed, and it is shown that the maximum input voltage is one fourth the diode Zener-voltage and is obtained when the pulse height is adjusted to one half the diode Zener-voltage. Because the negative pulse cannot entirely be eliminated, this ideal maximum input voltage cannot quite be obtained.

In connection with the biasing arrangement shown in Fig. 7, it may be noted that any circuit configuration that yields a volt-ampere characteristic like that shown in Fig. 11 will produce the desired results. The voltages e_1 and e_2 must be greater than twice the desired maximum output voltage. Hence the diodes and *R-C* (resistance-capacitance) bias networks may be replaced with diodes and fixed bias (Fig. 12), or by back-to-back diodes of equal or unequal Zener voltages as needed (Fig. 13), or simply by neon glow tubes (Fig. 14). The disadvantages of these arrangements compared to the *R-C* bias network include increased noise for the glow tubes, increased charging resistance for the back-to-back diodes, and circuit complication for the fixed bias. For these reasons, the *R-C* network is to be preferred for biasing the diode demodulator which has been described. This biasing method gives a noise level of about 3-mv (millivolt) rms with an output capacitance of 0.01 μf.

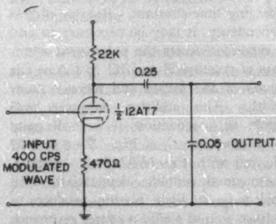


Fig. 10. Low-impedance driving source

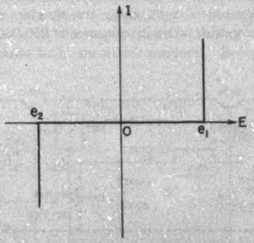


Fig. 11. Volt-ampere characteristic of switch element

The Magnetic-Amplifier Demodulator

The diagram, Fig. 15, shows a demodulator that uses the amplification properties of the reactors to obtain higher gain and to permit driving from sources of relatively high internal impedance. In this type of circuit the *R-C* network cannot be used for biasing because it is necessary that the bias voltage remain constant with varying pulse heights. The arrangement of Fig. 15 results in the lowest noise levels of the

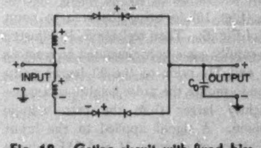


Fig. 12. Gating circuit with fixed bias

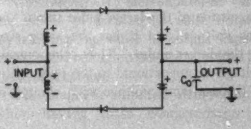


Fig. 13. Back-to-back diode gating circuit

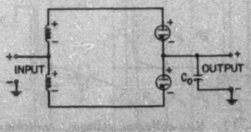
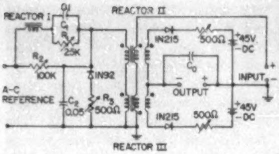


Fig. 14. Neon glow tube gating circuit



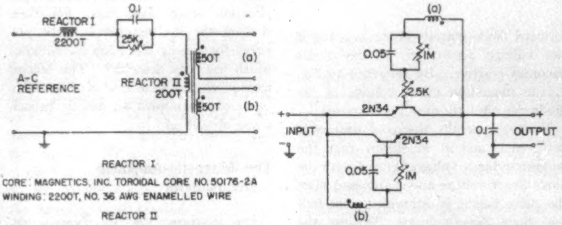
REACTOR I

CORE: MAGNETICS, INC. TOROIDAL CORE NO. 50176-2A
WINDING: 2200T, NO. 36 AWG ENAMELLED WIRE

REACTORS II AND III

CORE: MAGNETICS, INC. TOROIDAL CORE NO. 50153-1A
THREE WINDINGS: 400T, NO. 38 ENAMELLED WIRE

Fig. 15. Magnetic-amplifier demodulator



REACTOR I

CORE: MAGNETICS, INC. TOROIDAL CORE NO. 50176-2A
WINDING: 2200T, NO. 36 AWG ENAMELLED WIRE

REACTOR II

CORE: MAGNETICS, INC. TOROIDAL CORE NO. 50153-1A
PRIMARY WINDING: 200T, NO. 38 AWG ENAMELLED

WIRE. TWO SECONDARY WINDINGS: 50T NO. 38 AWG ENAMELLED WIRE

Fig. 16. Shunt-type transistor demodulator

various biasing methods previously shown.

The circuit operation is illustrated by Fig. 6. As noted before, pulse height may be varied by reducing the pulse area beyond point *A*. This reduction is best accomplished by adjusting resistor R_3 (Fig. 15) to bring the area to point *A* of Fig. 6. Then resistor R_2 is adjusted to supply some volt-time area to reactors II and III prior to the 90-degree point and to reduce the pulse heights to a value slightly larger than the bias voltage chosen. A signal applied to the input windings of reactors II and III reduces the pulse height of one of these reactors and increases the pulse height of the other. The output voltage is equal to the magnitude of the larger pulse minus the bias voltage, and it has the same sign as the larger pulse. Thus, the output voltage is proportional to the peak value of the a-c input voltage with about the same linearity, input impedance, and gain as are exhibited by a magnetic amplifier.

As in the case of the diode demodulator, the speed of response is dependent upon the pulse width as well as upon the charging time-constant. Because of this dependency, it may be necessary to add a capacitor across the pulse-input windings of reactors II and III to reduce the slopes of the pulses and increase their widths. This addition, however, will result in a reduction of over-all gain. With capacitor C_2 of Fig. 15 set at 0.02 μf, and with the *1N92* diode omitted, the magnetic amplifier demodulator with back-to-back diode biasing produces a voltage gain of 5 with a source resistance of 10,000 ohms. With output capacitance set at 0.01 μf, 1-cycle response is obtained. The noise level, however, is 70-mv rms.

By use of a matching transformer, a source with internal resistance of 250,000 ohms may be used. The resulting voltage gain is unity. Ideally, the desired turns for source matching could be wound on the core itself, but this is likely to prove impractical if small cores are used. The minimum number of turns that may be used for the pulse-input winding of the cores is governed by the cross-sectional area of the core, the properties of the core material, and the need for minimizing the magnetizing current. The minimum number of turns on the output winding is governed by the pulse height required. After these windings have been placed on the core, the remaining window area is often too small to accommodate the number of control winding turns desired.

If a fixed bias is used for the circuit shown in Fig. 15, and if the capacitor C_0 is set at 0.1 μf, a noise level of 20-mv rms is obtained. One-cycle response can be obtained with an output capacitance of 0.1 μf by setting capacitor C_2 at a value of 0.05 μf. The resulting voltage gain is 3.7 with a source resistance of 10,000 ohms. By use of fixed bias, the output capacitance can be increased without decreasing the speed of response. The loss of gain in this fixed-bias circuit results from the use of the *1N92* diode required to reduce the negative pulse height and to permit good linearity between ±18-volt output. The range of linearity may be extended by increasing the bias beyond the 45 volts used and eliminating the *1N92* rectifier.

When compared with the diode demodulator, the magnetic-amplifier demodulator suffers in that it is no more linear than the usual magnetic amplifier, while the diode demodulator is inherently linear over its range. In addition, the noise level is considerably higher. It should be noted that any point of the output loop may be grounded, because the output loop is isolated from the rest of the circuit by the reactor windings. An important requirement of this circuit is the use of matched cores for reactors II and III to ensure equal gain for in-phase and 180-degree out-of-phase input signals.

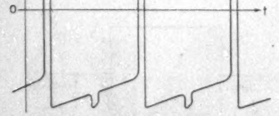
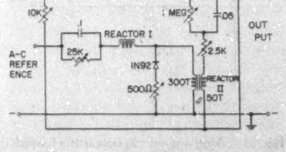


Fig. 17. Base-collector voltage: switching voltage wave form



REACTOR I

CORE: MAGNETICS, INC. TOROIDAL CORE NO. 50176-2A
WINDING: 2200T, NO. 36 AWG ENAMELLED WIRE

REACTOR II

CORE: MAGNETICS, INC. TOROIDAL CORE NO. 50153-1A
PRIMARY WINDING: 300T, NO. 38 AWG ENAMELLED WIRE
SECONDARY WINDING: 50T, NO. 38 AWG ENAMELLED WIRE

Fig. 18. Series-type transistor demodulator

The Shunt-Type Transistor Demodulator

The shunt-type transistor demodulator shown in Fig. 16 provides 1-cycle response at low noise levels. Its operation is similar to that of the preceding demodulators. During the interval between the gating pulses, the *R-C* network provides sufficient bias to keep the transistors shut off. Because the transistors will conduct unless both the emitter and collector are negative with respect to the base (for p-n-p transistors),[3] the cut-off bias voltage must be greater than twice the maximum output voltage or, equivalently, greater than the maximum peak-to-peak input voltage. The maximum bias that can be used is fixed by the maximum inverse-voltage rating of the transistors, about 40 to 50 volts for type *2N34* transistors. Hence, the maximum output voltage is approximately ±20 volts. Reference to Fig. 17, which shows the base-to-collector voltage applied to the transistor, reveals that the voltage that turns on the transistor during the pulses is dependent upon the d-c level of the wave form. The d-c level is set to provide correct cutoff bias. The turns ratio of reactor II (Fig. 16) therefore must be chosen in such a way that both the on and off base-to-collector voltages are sufficient. To prevent excessive power dissipation and destruction of the transistor, the current used to turn on the transistor should be greater than one half the maximum emitter current when the common-collector connection is used.[3] When these conditions have been met, however, the inherent efficiency of the switching operation permits the transistors to control much greater power than they could if operated as linear amplifiers.

The switching-voltage wave form shown in Fig. 17 contains positive and negative pulses, an alternating reference-voltage component, and a decay voltage from the *R-C* bias network. All of these voltages feed through to, or are picked up by, the output circuit, but the symmetry of the circuit assures cancellation of the larger part of these voltages. Only the mismatch between the two transistors results in feed-through to the output. The equal parts of these feed-through voltages result in circulating currents within the demodulator itself, and do not affect the output voltage. Rough balance of the transistors may be obtained by selection, and may be refined by adjustment of the base currents. With output capacitance of 0.02 μf, the resulting noise output is about 3-mv rms. Maximum output is ±25 volts.

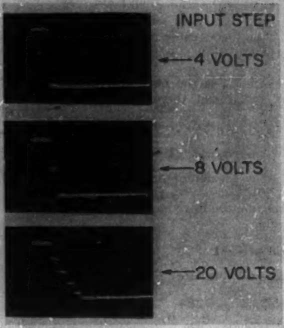


Fig. 19. Step response of series demodulator

The Series-Type Transistor Demodulator

The series-type transistor demodulator shown in Fig. 18 is very similar in operation to the shunt-type demodulator except that it requires only about one volt of bias to shut it off satisfactorily. This feature is a consequence of the series connection of the transistors, in that the output voltage itself causes one or the other of the two emitters to be negative with respect to the base, while both collectors are held negative by the bias voltage. Hence, at all times except during the pulse, one transistor is shut off while the other may be off or on depending on the magnitude of the input signal. During the pulse both transistors conduct. Because the pickup and feed-through voltages occur during the time between pulses when the circuit may be unsymmetrical (one transistor on, one off), cancellation of the noise voltages is not as complete as it is in the case of the shunt-type demodulator. In addition, the series connection of the transistors results in a charging resistance approximately four times as great as that of the shunt-type demodulator. This increase in charging resistance must be matched by a corresponding decrease in output capacitance if speed of response is to be maintained. The resulting output noise is about 14-mv rms as compared to 3-mv rms for the shunt connection. Maximum output is ±30 volts for type *2N34* transistors.

The series demodulator has a peculiarity, however, which enables it to give one cycle response to small input steps, while requiring much more time for larger steps. Reference to Fig. 19 shows that the output voltage response to a step input tends to rise by a fixed amount during each pulse, independent of the size of the input step. The final value of output is reached on a nearly linear rise instead of following the usual exponential rise. The output capacitor appears to be charged during the pulses from a current source of constant value. This means that the current source is independent of the size of the input step. Hence, for steps smaller than the amount that the voltage on the output capacitor changes during each pulse, 1-cycle response is obtained. This feature allows a considerable increase in the value of output capacitance with a corresponding decrease of output noise, if 1-cycle response is required only for small steps.

Conclusions

The demodulators described provide 1-cycle response to step inputs, at varying values of zero signal noise output. By varying the value of output capacitance, noise reduction may be sacrificed for speed of response.

The magnetic-amplifier demodulator, while having these characteristics in

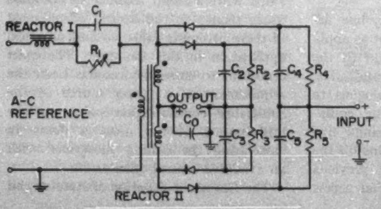
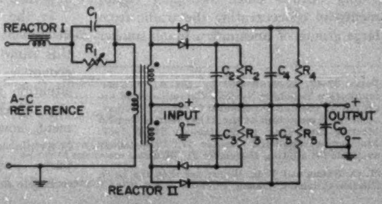


Fig. 20 (left). Modulator

Fig. 21 (right). Quadrature rejector



common with the other demodulators described, has the additional characteristic of a high input impedance. By lowering the input impedance, the voltage gain may be increased.

The diode and transistor demodulator circuits also may be adapted for use as a modulator or as a quadrature-rejection circuit. Adaptions of a diode-demodulator circuit are seen in Figs. 20 and 21.

Appendix. Maximum Output Voltage

If it is assumed that the diodes have zero forward impedance, and that C_0 is very much less than C_1 and C_2, it is possible to determine the theoretical maximum signal handling capability of the demodulator circuit in terms of the diode reverse breakdown or Zener voltage, E_z. If it is assumed that the input signal is of the form $E_i \sin \omega t$, and is of the same frequency and phase as the reference voltage, the change on capacitor C_0 has the polarity shown in Fig. 7. Equations can be written for the voltage across each diode during the interval between pulses as follows:

$$-E_{D1} = E_{C1} + E_0 - E_i \sin \omega t \qquad (1)$$

$$-E_{D2} = E_{C2} - E_0 + E_i \sin \omega t \qquad (2)$$

where $-E_{D1}$ and $-E_{D2}$ represent the voltages across the diodes in the reverse direction. From the assumed conditions it can also be said that

$$|E_0| = |E_i|$$

and

$$|E_{C1}| = |E_{C2}| = |E_p|$$

where E_p is the pulse height. Equations 1 and 2 can therefore be rewritten

$$-E_{D1} = E_p + E_i(1 - \sin \omega t)$$

$$-E_{D2} = E_p - E_i(1 - \sin \omega t)$$

The maximum and minimum values of $-E_{D1}$ and $-E_{D2}$ respectively occur when the input voltage has its maximum negative value

$$\sin wt = -1$$

These values are

$$-E_{D1} = E_p + 2E_i$$

$$-E_{D2} = E_p - 2E_i$$

The circuit no longer operates properly if, during the interval between pulses, either of the diodes conducts. Conduction occurs when either

$$-E_{D1} \geq E_z$$

or

$$-E_{D2} \leq 0$$

Therefore for proper operation the maximum values are limited to the range specified by the following inequalities:

$$-E_{D1} = E_p + 2E_i \leq E_z$$

$$-E_{D2} = E_p - 2E_i \geq 0$$

These can be rewritten as

$$E_i \leq \frac{1}{2}(E_z - E_p) \qquad (3)$$

$$E_i \leq \frac{1}{2} E_p \qquad (4)$$

The maximum value of E_i which satisfies both equation 3 and equation 4 occurs when E_p is chosen as

$$E_p = \frac{1}{2} E_z$$

The equation for the maximum value of E_i is therefore

$$E_i = \frac{1}{4} E_z$$

By a similar analysis it may be shown that this same value of E_p also provides the same maximum signal-handling capability when the phase of the input signal is reversed.

References

1. High Performance Demodulator for Servomechanisms, K. E. Schreiner. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. II, Oct. 3–5, 1946, pp. 393–403.

2. Application of a Magnetic Amplifier to a High Performance Instrument Servo, Paul R. Johannessen. *Convention Record*, Institute of Radio Engineers, pt. IV, New York, N. Y., Mar. 21–24, 1955, pp. 15–22.

3. Junction Transistors Used as Switches, R. L. Bright. *AIEE Transactions*, vol. 74, pt. I, Mar. 1955, pp. 111–21.

The Automatic Line Concentrator and Subscriber Telephone Applications

H. G. EVERS
ASSOCIATE MEMBER AIEE

J. R. WYLIE
AFFILIATE MEMBER AIEE

WHERE a large number of telephone subscriber lines are to be served, it is possible to show statistically that only a small percentage of them will require service at a given interval. Telephone operating companies have used this fact to economic advantage in providing central office switching equipment, by concentrating the traffic from large groups of lines into a much smaller number of primary switching trunks. Generally speaking, 100 lines may normally be served with only from 10 to 15 primary trunks without degrading service beyond the point where one call in one hundred will encounter an appreciable delay, so that the economy of this approach is obvious.

In the United States, party-line development was an early attempt to apply this same statistical advantage to the economic use of outside plant. The situation is not precisely analogous to that of central office switching equipment, however, since only a single line is available to all of the stations on the party line, whereas in the previous example each of the 100 lines has access to each of the primary switching trunks as required. Furthermore, the calling habits of subscribers who can conveniently be partied together are likely, because of geographic and economic similarities, to raise the incidence of delay far over the statistical prediction. Although the quality of service thus deteriorates rapidly as the number of stations per line is increased, the party line has become the accepted American method of effecting economy in outside plant. While it is true that from time to time various mechanical devices have been proposed which would achieve more uniform concentration by associating a substantial number of short subscriber loops with a small number of common usage trunks to the central office, most of these attempts failed to find practical application in this country. The chief objection to these devices has been the requirement of a power source at the subscriber concentration point, or the introduction of too great a delay in associating the calling subscriber with an available central office trunk.

The European situation of state-owned

Paper 57-47, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 17, 1956; made available for printing November 21, 1956.

H. G. Evers and J. R. Wylie are with the Leich Electric Company, Genoa, Ill.

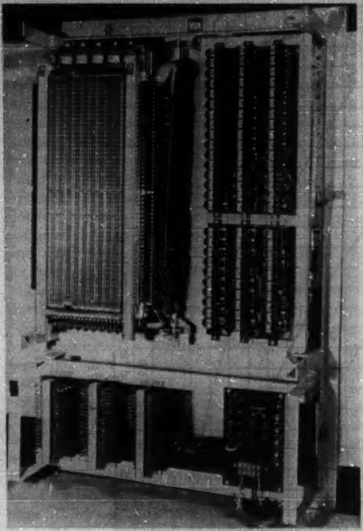


Fig. 1 (left). Line concentrator unit of central office

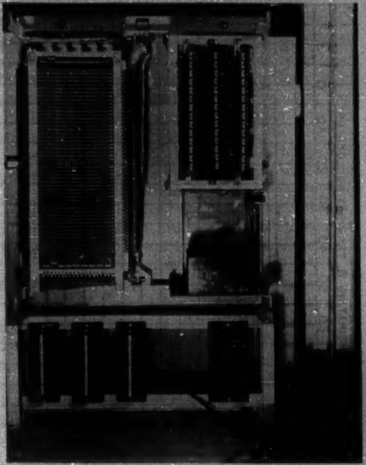


Fig. 2 (right). Line concentrator unit of sub wire-center

communication facilities and a consequent or inborn insistence upon privacy of conversation has effectively ruled out the party line as a method of effecting economy in plant facilities on the continent. The need for an alternate solution has intensified development of subscriber line concentrators to provide service where a number of subscribers are grouped together at some distance from the central office, as in an apartment building or within a rural community. About 10 years ago, Christoph Gfeller AG, of Bern, Switzerland, introduced such a concentrator using in conjunction with a group of control relays, a form of crossbar switch which because of its low-power requirements and rapid action offers promise of overcoming American objections to earlier mechanical concentration units of domestic or foreign manufacture.

During the past 2 years, Leich Electric Company, in co-operation with its affiliated operating companies within the General Telephone System, has been conducting field trials on concentrators of Gfeller manufacture, similar to those shown in Figs. 1 and 2, in an effort to assess their economic value and reliability as a means of reducing outside plant investment. Results of these tests have been most encouraging, with prompt and reliable operation obtained as well as indicated increases in outside plant efficiency of from 35 per cent to more than 100 per cent, depending upon the application.

The basic application of these devices, shown in Fig. 3, is to reduce the number of conductors required over a greater part of the distance from the central office to a group of subscribers, by establishing a sub wire-center in their immediate geographical area, while retaining an individual pair per line within the central office, and between this center and the stations. A remote concentrator unit, about 2 by 3 feet in size and 9 inches deep, is located within a suitable weatherproof enclosure at this distribution point. It is connected over a group of 9 to 18 traffic trunks and several control trunks to a second unit, the same size as the first and electrically its mirror image, located at the central office. When not more than 12 traffic trunks are needed, two control trunks suffice to register selections and to supply power. The additional selections required when more than 12 traffic trunks are provided necessitate the addition of a third control trunk.

All power is furnished from the central office. Direct current used for trunk and line selections is fed over one of the control conductors to a large electrolytic capacitor (3,500 microfarads) which serves the subscriber unit as a power reservoir, maintaining the necessary potential on the feeder wire during momentary periods of peak power requirement as the apparatus prepares for each selection.

The heart of the device is the Trachsel selector switch, a unique application of the crossbar principle in which the ultimate connection is maintained wholly by the interaction of gravity and spring tension, with no electric energization required once the contact has been established. The switch is vertically mounted, with horizontal selection members equal to the number of lines to be served, and a vertical bar for each

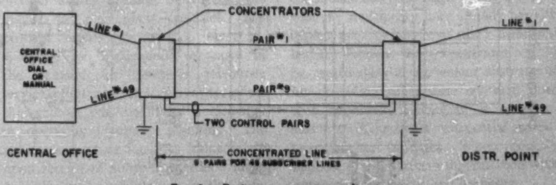


Fig. 3. Basic concentrator application

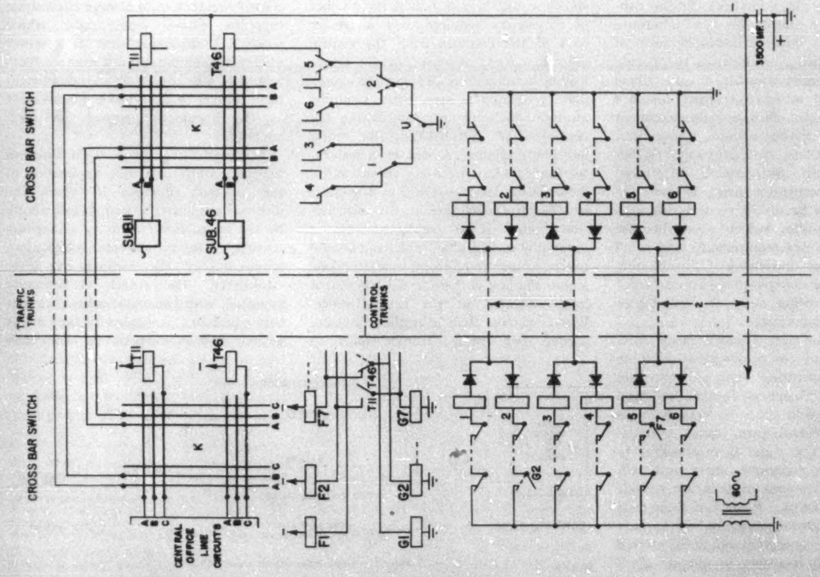

CROSS BAR SWITCH
TRAFFIC TRUNKS
CROSS BAR SWITCH
CENTRAL OFFICE LINE CIRCUITS
A B C
K
T11
T46
F1
F2
F7
G1
G2
G7
CONTROL TRUNKS
SUB.11
SUB.46
U
60~
3500 MF
1 2 3 4 5 6

Fig. 4. Principle of operation of crossbar line concentrator

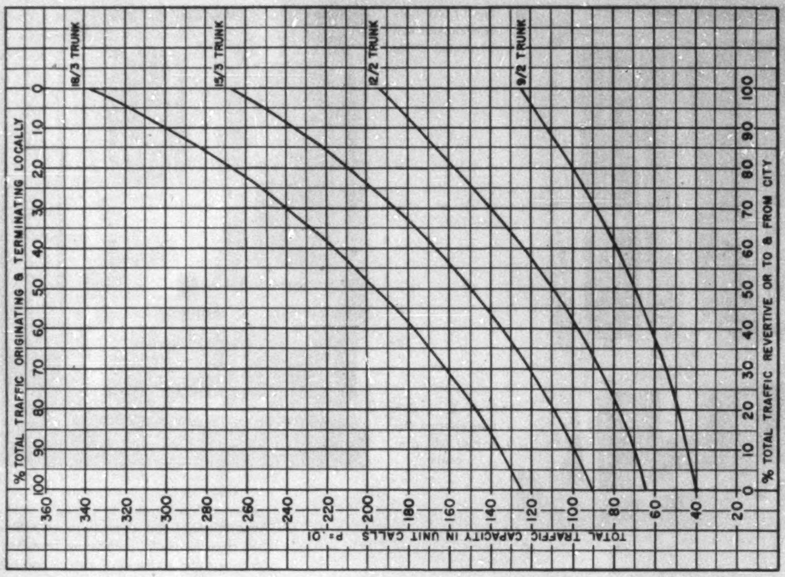


Fig. 5. Traffic capability curves for line concentrator

traffic trunk. In practice, a vertical bar is lifted electromagnetically to preselect a traffic trunk and is held in this position, with reduced magnet energization fed over the traffic trunk from the central office. On the initial service demand for a particular line, the corresponding horizontal member is moved aside under the control of a relay individual to that line. At the subscriber unit each such relay is operated by loop closure; in the exchange unit each is controlled by the sleeve lead, or test wire. This brief sidewise movement of the horizontal member places a set of multiple pin conductors in the path of travel of each of the vertical bars. The holding circuit for the one which has been preselected is then interrupted, permitting that bar to fall back into place. As it does so, all but one of its multiple sets of wipers, located at each intersection with the horizontal members, return to position without interference. At the operated horizontal member, however, the wipers engage a set of the pin conductors and force them aside. This movement is also imparted to a set of cutoff contacts associated with the given line, opening them to disconnect the control relay from the line. The pin conductors for that line thus return to their rest position except for those just engaged by the released vertical bar, the wipers of which hold them firmly and maintain a connection as long as a call is in progress, and thereafter until that traffic trunk is selected in rotation in preparation for a new call. When the bar is raised at that time, the pin conductors from the previous connection are disengaged from the wipers of the bar and thereby released.

In most models of this concentrator, operation of the vertical and horizontal members of the crossbar switch is controlled by a group of six marker relays. These relays are connected, two at each unit, to each of three control conductors in a series-parallel arrangement, as shown in Fig. 4. Commercial alternating current supplied at the central office is fed through rectifiers so that one relay of each pair in each unit is energized during the positive half-cycle, the other during the negative half-cycle. Circuit constants are such that each relay remains operated during the half-cycle when it is not energized. Selections are transmitted to the distant unit by opening at the sending end any combination of the six rectified branches, thus causing release not only of a relay in each affected branch at the sending unit, but of the corresponding relays in the distant unit. In this manner, identical selections are made in both units when a subscriber initiates a call, or when a call from the central office is made to a subscriber served by the concentrator.

The 6-relay marker described will suffice for a 50-line unit with 12 traffic trunks, since 63 marking combinations (2^6-1) are provided, 50 for line selections, 12 for trunk selections, and one spare. To meet varying service needs the concentrator units are available with 19 lines and 9 trunks, 29 lines and 12 trunks, 49 lines and 9 trunks, or 49 lines and 18 trunks. In each unit one additional line position is reserved for test purposes.

A group of subscribers served by such a concentrator is thus provided with a small number of traffic trunks, any one of which can be used by any of the affected lines. Each trunk is used in rotation as a new connection is required, if it is idle at that time. Since a selection, once established, is not disturbed when the call is completed, but only at some period thereafter when that trunk is again reached in rotation and prepared for a new selection, it is possible for a subscriber to make several additional calls without causing any further operation of the concentrator. Each call will be handled over the same trunk as the first, to which the subscriber's line remains connected until, at some period between calls, this trunk is required in preparation for the next call to be placed. The next such call could, of course, be placed by the same subscriber, who would again be connected to the same trunk as before.

The connection established by the concentrator is direct and metallic, with absolutely no repeat coils, impulse relays, or foreign potentials introduced into the circuit. Since no local battery or ringing generator is available at the subscriber unit, calls between lines served by the same concentrator require two trunks, as they must be switched, rung, and fed battery from the central office equipment.

The outstanding example of the economic value of these devices, at present, is their ability to meet unexpected demands for new plant facilities. The postwar housing boom has been responsible for tremendous increases in the demand for residence telephone service, often concentrated in areas where this growth had not been anticipated in operating company planning. The ability to provide facilities immediately upon receipt of subscriber orders, made possible by a concentrator in the area, is of distinct value in the company's customer relations.

While the unit may be permanently installed advantageously in some cases, in many instances a single concentrator may be used in several successive locations as a stop-gap solution until permanent cable relief can be provided. The saving effected by deferring carrying charges on such cable plant until a time when the expansion can be incorporated in orderly fashion into the company's financial planning will often amortize the acquisition cost of the concentrator on first application.

Such emergency relief, however, is not the sole purpose of this equipment. A group of multiparty lines can readily be concentrator-adapted to provide upgraded service to the affected subscribers at a moderate investment in facilities. Such action is profitable not only in terms of the revenue increase which this upgrading will provide, and the accompanying subscriber goodwill because of the improved service, but also by virtue of the increased toll traffic which the more ready availability of switching facilities will generate on the concentrated lines.

Of prime importance in considering the use of concentrators is to match the application to the capability of the device. Field tests have indicated the advisability of a thorough traffic study to determine not only the incidence and duration of calls from the lines in question, but also the part of this traffic which both originates and terminates within the line group. As may be seen in the traffic curves of Fig. 5, the capacity of the unit is affected by the relative proportion of such calls, as they require two traffic trunks for completion. The choice of concentrator will thus depend upon traffic potential, suitably weighted in consideration of this factor.

As has been shown, the impetus for concentrator development arose from a situation in which party-line service was precluded. These two methods of effecting plant economy are therefore essentially alternate rather than supplementary. Accordingly, the concentrator should not be thought of as a means whereby multiparty lines already heavily loaded may be made to carry increased traffic. Rather, the concentrator should be regarded as a convenient instrument for providing subscribers on such lines with improved service at a commensurate tariff, without substantial investment in plant facilities. When properly applied, it will perform this function with impressive flexibility and economy.

On the Input Power of the Half-Wave Magnetic Amplifier Circuit

T. KIKUCHI
ASSOCIATE MEMBER AIEE

K. MURAKAMI
NONMEMBER AIEE

Synopsis: The mechanism of amplification of the self-saturating magnetic amplifier in general consists of its flux control of the magnetic core. To control the flux level in the core, the core loss must be supplied from the control circuit during the control or reset half-cycle of the specific core. Therefore the correlation of input power and reset core loss is very important. (Here the reset core loss means the portion of the core loss consumed during the control or reset half-cycle of the core.) In the case of the full-wave magnetic amplifier circuit, this relation has been discussed in the literature;[1] however, the case of the half-wave magnetic amplifier has not received as detailed consideration. In this paper the authors demonstrate the correlation between the input power and the reset core loss utilizing a simplified method of circuit analysis. Experimental procedure derived from the simplified analysis verify the analytical conclusions. The existence of an optimum condition is shown. The optimum requires the resistance of the control voltage source and the effective resistance of the magnetic core during the resetting interval to be matched. This condition permits achievement of minimum input power.

Assumptions

THE circuit phenomena of the half-wave magnetic amplifier are not so simple even in the case of the ideal conditions.[2] To grasp the chief features of the half-wave magnetic amplifier without missing its important points, it is best to simplify the several circuital conditions and facilitate the analysis.

SIMPLIFIED CIRCUIT

In place of the usual half-wave magnetic amplifier circuit shown in Fig. 1(A), let us consider the circuit of Fig. 1(B) with a synchronous switch in its output circuit which closes only during the positive half-cycle of the a-c supply voltage. The phenomena in each of the circuits are almost the same when the magnitude of the control voltage E_c is small; the phenomena are different when the control voltage becomes large.

APPROXIMATE REPRESENTATION OF CORE CHARACTERISTICS

The core is assumed to have a rectangular hysteresis loop such as that which the grain-oriented ferronickel alloys exhibit. To make the simplified analysis successful, the appropriate approximate representation of the magnetic characteristics is very important. The following approximation for the descending branch of the dynamic characteristics of the core is employed; see Appendix I.

$$\left.\begin{array}{l} N_c i_c = N_c I_\infty - k\, d\phi/dt \\ \text{for } d\phi/dt<0 \text{ and } |\phi|<\phi_S \end{array}\right\} \qquad (1)$$

where N_c is the number of turns of the control winding and ϕ and ϕ_S are respectively the instantaneous value of the core flux and its saturation value. The current i_c is the instantaneous value during the control interval and I_∞ is the constant deduced from the core characteristics as explained in Appendix I.

Input Power of Half-Wave Magnetic Amplifier and Optimum Condition of Control Circuit

In the circuit of Fig. 1(B), consider the situation during the control half-cycle, then the following relation is satisfied

$$E_c = R_c i_c - N\, d\phi/dt \qquad (2)$$

And using equation 1 together with equation 2, the value of $d\phi/dt$ will be proved to be constant during the control half-cycle, and consequently the value i_c will also be a constant value I_0. Let $\Delta\phi$ represent the amount of flux which is reset during the control half-cycle, then the rate of flux change $d\phi/dt$ during this interval will be equal to $-2f\Delta\phi$. Then equation 2 can be simplified as

$$E_c = R_c I_0 + 2fN_c\Delta\phi \qquad (3)$$

Let I_c = the average value of the control current, and P_c, the input power, then

$$I_c = E_c/R_c \qquad (4)$$

and using the relation of equation 3, the input power can be expressed by the following equation

$$P_c = (R_c I_0 + 2fN_c\Delta\phi)^2/R_c \qquad (5)$$

Using the relation in equation 5, the optimum value of the control circuit resistance, which minimizes P_c for the specified amount of $\Delta\phi$, can be easily calculated

$$R_{c0} = 2fN_c\Delta\phi/I_0 \qquad (6)$$

In equation 6, $2f\,N_c\Delta\phi$ may be considered as the counter electromotive force of the magnetic core which appears across the control winding during the control half-cycle, and I_0 is the exciting current during that interval. The right-hand side of equation 6 represents the equivalent resistance of the magnetic core seen from the control winding during that interval. That is, the relation implies that the equivalent resistance of the core is matched to the control circuit resistance. It should be noted that this equivalent resistance is not constant but is the linear fractional function of the amount of reset flux, $\Delta\phi$ for this case, and in general it is a nonlinear function of $\Delta\phi$.

The minimum input power P_{c0}, required to realize the specified amount of reset flux $\Delta\phi$, can be calculated from equations 5 and 6, giving the results as follows:

$$P_{c0} = 8\times f\times N_c\times I_0\times\Delta\phi \qquad (7)$$

On the other hand, the reset core loss P_r consumed in the core can be easily calculated as

$$P_r = f\times N_c\times I_0\times\Delta\phi \qquad (8)$$

From equations 8 and 9

$$P_{c0} = 8P_r \qquad (9)$$

Equation 9 means that the control power of the half-wave magnetic amplifier is at least eight times the reset core loss under the idealized conditions. This is just the reason for the low power gain of the half-wave magnetic amplifier. It should be noted, however, that the basic half-wave circuit may be improved by suitable modification.[3,4]

CHOICE OF CONTROL WINDING TO MATCH GIVEN CONTROL VOLTAGE SOURCE

In the following, consider the problem of how to realize the maximum amount of reset flux $\Delta\phi$ by choosing the appropriate number of turns of the control winding when the control voltage E_c and the control circuit resistance R_c are given.

In this case, the rate of flux change $d\phi/dt$ during the control half-cycle will take various values when the number of turns

---

Paper 57-5, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted July 3, 1956; made available for printing October 18, 1956.

T. KIKUCHI is with the Microwave Research Institute of Polytechnic Institute of Brooklyn, on leave from Tohoku University, Sendai, Japan; and K. MURAKAMI is with Tohoku University, Sendai, Japan.

The authors wish to thank Profs. Ernst Weber and E. J. Smith of the Polytechnic Institute of Brooklyn for presenting the opportunity of publishing this paper; Joseph Antin of the Polytechnic Institute of Brooklyn for his help with English expression, and his kind comments; and Prof. Y. Watanabe and K. Fukushima of Tohoku University for their encouragement and guidance.

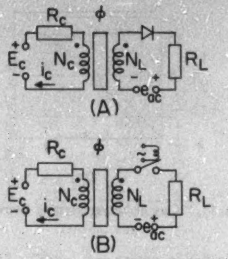


Fig. 1. Basic half-wave amplifier circuit

of the control winding changes. Therefore, to represent the descending portion of the hysteresis loop, it is necessary to return again to equation 1. Starting from equations 1 and 2, the following relation is obtained:

$$N_c(E_c - R_c I_\infty) = -(R_c k + N_c^2)d\phi/dt \qquad (10)$$

Integrating this equation during the control half-cycle gives

$$2f\Delta\phi = N_c(E_c - R_c I_\infty)/(R_c k + N_c^2) \qquad (11)$$

and the maximum value of reset flux $(\Delta\phi)_{max}$ can be calculated, considering E_c and R_c as constants, and the number of turns of control winding to realize this condition as N_{c0}

$$2f(\Delta\phi)_{max} = E_c/2N_{c0} \qquad (12)$$

This equation means that one half of the control voltage is applied to reset the core and the remaining half is consumed in the circuit resistance. This is, again, the condition of matched impedances during the control half-cycle.

To choose the appropriate number of turns of the control winding, the equivalent resistance of the core seen from the control winding should be equal to the control circuit resistance R_c. Then the optimum reset power which is consumed in the core during the control half-cycle is

$$P_r = E_c^2/8R_c \qquad (13)$$

On the other hand, the relation between the reset core loss P_r and the amount of reset flux $\Delta\phi$ can be found by the method explained in the paper; therefore the optimum number of turns is given by equation 12 and may be written as

$$N_{c0} = E_c/4f\Delta\phi \qquad (14)$$

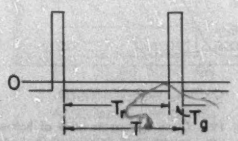


Fig. 2. Pulsed supply voltage

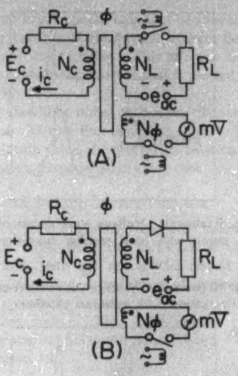


Fig. 3. Basic circuits for experimental verification

Control Power When Length of Control Period Differs from Half-Cycle

In the case of the half-wave magnetic amplifiers utilizing a pulsed supply voltage,[5] matters will differ to some extent from the foregoing considerations.

Assume the wave form of the supply voltage shown in Fig. 2, and represent the length of the control interval by T_r and the output or gating interval by T_g. Assume that the output circuit closes only during the interval T_g in the basic circuit of Fig. 1(B).

The fundamental equation during the control interval can be deduced as

$$E_c = R_c I_0 + N_c \Delta\phi/T_r \qquad (15)$$

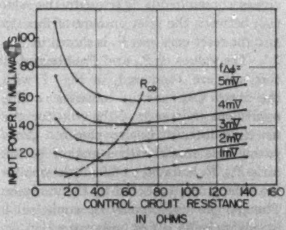


Fig. 4. Relations between control circuit resistance and input power

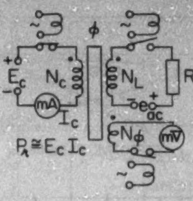


Fig. 5. Circuit for measuring reset core loss

and the input power is

$$P_c = (R_c I_0 + N_c \Delta\phi/T_r)^2/R_c \qquad (16)$$

From this the minimum input power for the specified amount of reset flux is

$$P_{c0} = 4 I_0 N_c \Delta\phi/T_r \qquad (17)$$

whereas the reset core loss necessary to give the reset amount of flux $\Delta\phi$ is

$$P_r = \frac{1}{T} N_c \Delta\phi \times I_0 \qquad (18)$$

$$\text{and } \frac{P_{c0}}{P_r} = 4\frac{T}{T_r} \qquad (19)$$

In the limit, the width of the pulse becomes very small

$$\lim_{T_r \to T} \frac{P_{c0}}{P_r} = 4 \qquad (20)$$

Experimental Verification

Experimental verification of the simplified analysis is directed toward the confirmation of two basic points. The first point is the validity of the approximation of the dynamic characteristics of the magnetic core; the second concerns the

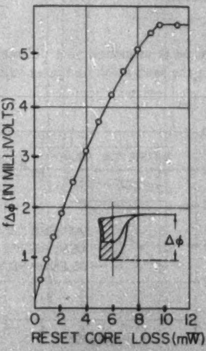


Fig. 6. Relation between reset core loss and amount of reset flux

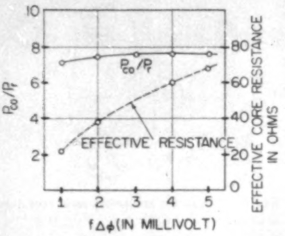


Fig. 7. Experimental value of the ratio P_{co}/P_r

usefulness of the analysis in the case of the usual half-wave magnetic amplifier circuit. Experiments are performed using the a-c supply voltage source of 50 cycles per second, and a magnetic core of grain-oriented 50 per cent ferronickel alloy core.

EXPERIMENT UNDER SIMPLIFIED CONDITION

First, the validity of the simplified analysis is checked by an experiment, under conditions faithful to the analytical requirements. In the experiments, the amount of reset flux $\Delta\phi$ is adopted as a basic parameter. To measure the reset flux, vibratory contact rectifiers are used with an average sensitive d-c millivolt meter as a flux voltmeter.[6-8] The schematic diagrams are shown in Fig. 3. Represent by N_ϕ the number of turns used for measuring $\Delta\phi$ and V_ϕ, the rectified average reading of the d-c millivolt meter, then by choosing the proper length of the closing interval and the phase angle of the vibratory rectifiers, the amount of the reset flux $\Delta\phi$ can be easily measured with the use of the following relation:

$$f\Delta\phi = V_\phi/N_\phi \qquad (21)$$

where $\Delta\phi$ is measured as a difference between the saturation and reset values.

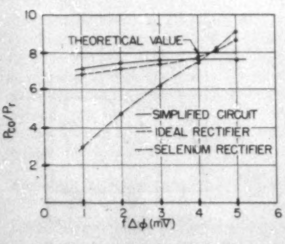


Fig. 8. Comparison of the results of several circuits

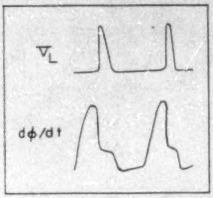
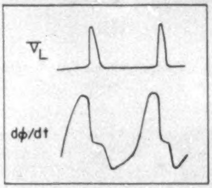


Fig. 9 (above). Voltage wave shapes of the circuit with ideal rectifier condition

Fig. 10 (below). Voltage wave shapes of the circuit with selenium rectifier



In the circuit of Fig. 3(A), the magnitude of a-c supply voltage is made large enough to maintain the same peak value of the gate magnetizing ampere turns, as in the case of measuring the control magnetization curve.[6,9] The gate circuit is closed during the positive half-cycle of the a-c supply voltage. Under these conditions, the length of the control interval can be kept to 1/2 cycle just as in the case of the simplified analysis. With the reset amount of the flux $\Delta\phi$ kept constant, as a parameter, control circuit resistance is plotted against the input power, and the family of curves shown in Fig. 4 is obtained. As expected from the analysis, the minimum control power P_{co} corresponding to each value of $\Delta\phi$ can be found easily from the figure.

To compare the results with the simplified analysis, it is necessary to know the reset core loss. To measure this, the circuit shown in Fig. 5 is used; the relation between the reset amount of flux $\Delta\phi$ and the reset core loss P_r is shown in Fig. 6.[7] The values of P_{co} and P_r obtained in this way are compared in Fig. 7 with the reset flux as the abscissa. The equivalent resistance of the core for the resetting interval is also shown. Comparing this result with equation 9, the ratio P_{co}/P_r is found to be a little smaller than the value predicted by the analysis. Though the assumption of equation 1 is valid only as a first approximation, it will be adequate in practice since the optimum relation obtained experimentally differs only slightly from equation 9.

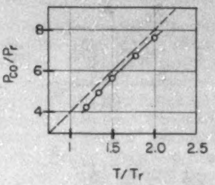


Fig. 11. Case of longer resetting interval T_r

To examine the effect of the parallelogramic hysteresis loop, calculations are made in Appendix II; the results calculated are almost the same as those obtained in the case of the simplified analysis considered previously. However, the computations do not completely explain the experimental results. Further detailed study will be necessary to clarify this discrepancy.

HALF-WAVE MAGNETIC AMPLIFIER WITH IDEAL RECTIFIER

As an intermediate stage in comparing the simplified analysis with the practical circuit, the case of the magnetic amplifier with an ideal rectifier is studied experimentally. The ideal rectifier can be realized experimentally by adopting the vibrating contact rectifiers in the gating circuit as shown in Fig. 3(A) and adjusting the closing interval of the gating circuit so as to close the contacts only during the time interval when a positive voltage is applied to the load resistance R_L. (The rectifier adjustment may be readily accomplished by observing the wave shape of the induced voltage.) Under this condition, measurements required to obtain results analogous to Fig. 4 are made,

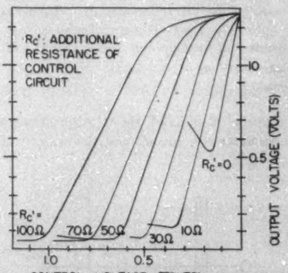


Fig. 12. Control characteristics of half-wave magnetic amplifier of various control circuit resistances

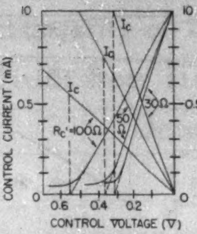


Fig. 13. Control characteristics of half-wave magnetic amplifier shifted with bias

and the results compared in Fig. 8. The small discrepancies compared with the former experiment are attributed to the difference in the reset voltage wave forms. That is, in the former the flux is almost reset by the applied rectangular voltage, whereas in the latter a trapezoidal voltage is applied to the core due to the back-firing.[9] This results in the slight difference in the dynamic behavior of the core, and the reset core loss will be decreased in this case for the region of small control voltage. Whereas, when the control voltage becomes larger, by the phenomenon of back-firing, the necessary control voltage to reset the specified amount of core flux becomes considerably larger in the case of the ideal rectifier condition than in the case of the simplified circuit condition. In Fig. 9 the wave form of the induced voltage of the core is shown for the condition $\Delta\phi=4mV/f$.

HALF-WAVE MAGNETIC AMPLIFIER WITH SELENIUM RECTIFIER

Similar experiments are done with the use of selenium rectifiers in the output circuit, as shown in Fig. 3(B). The results are also compared in Fig. 8. The wave form of the induced electromotive force during the resetting interval is rather complicated in this case, as shown in Fig. 10. In the figure the reset amount of flux $\Delta\phi$ is also taken at $4mV/f$. The control characteristics of the amplifier become less steep because of the inverse current flow through the rectifier. This effect is also noticeable in the relation that is plotted in Fig. 8.

It should be noted that when the reset amount of core flux is almost equal to its maximum value, the theoretical relation of equation 9 is satisfied in every case.

CASE OF PULSED SUPPLY VOLTAGE

The case where the resetting interval T_r is longer than the half-cycle is also checked, with the use of the circuit of Fig. 3(A), and substitution of a direct-voltage source in place of an alternating voltage source. Fig. 11 compares the experimental data with those predicted by equation 19.

MAGNETIC AMPLIFIER WITH BIAS WINDING

The optimum condition of the control circuit is also checked experimentally for the case of the half-wave magnetic amplifier with a constrained bias winding. This time, with the use of a magnetic core of Mo-Permalloy and measurement of the usual control characteristics by varying the control circuit resistance as a parameter, the group of curves shown in Figs. 12 and 13 is obtained. Fig. 12 shows the characteristics with no bias current. In Fig. 13, the bias ampere-turns in a constrained excitation circuit are chosen to give one volt of output voltage when the control voltage is zero at any value of the control circuit resistance. Approximating the characteristics in Fig. 13 by straight lines tangential to them, the amount of input power necessary to give rise to the variation of one volt of output voltage is calculated. Fig. 14 is the relation between the control circuit resistance and the input power thus obtained. The optimum value of the control circuit resistance is about 50 ohms and this also is almost equal to the equivalent input resistance (measured by the use of the circuit of Fig. 5 with a constrained bias of 0.25 ampere turn) of the core during the resetting interval.

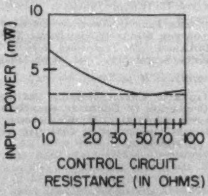


Fig. 14. Existence of optimum input circuit resistance

Conclusions

The relation of the reset core loss to the optimum input power in the case of the basic half-wave magnetic amplifier is made clear by the simplified analysis and experiments. The existence of an optimum control circuit condition which requires the impedance of the control circuit to be matched to the effective resistance presented by the core during the reset interval is demonstrated.

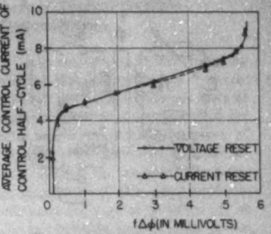


Fig. 15. Half-cycle average control current for two extreme conditions

Appendix I. Approximate Representation of Core Characteristics

The control circuit resistance utilized in the half-wave magnetic amplifier treated in this paper is low enough for the mode of flux reset to be regarded as approximately the voltage reset. This may be verified by observing the wave form of the induced voltage during the resetting interval.

In the case of direct voltage reset, which can be realized with the use of the circuit of Fig. 5, the average control current during the resetting half-cycle I_c can be easily measured. I_c will correspond to the average dynamic coercive force of the core. The relation between I_c and the amount of reset flux $\Delta\phi$ coincides almost perfectly with the control magnetization curve in the case of the grain-oriented 50 per cent ferronickel alloy; see Fig. 15. Therefore, with the use of the control magnetization curve,[5,9] the group of the approximate representations for the descending branch of the dynamic minor loops for specified value of $\Delta\phi$ can be readily obtained, as shown in Fig. 16. That is, the instantaneous value of the con-

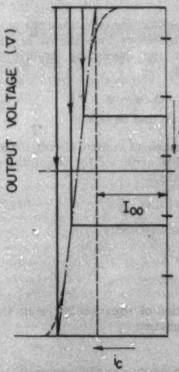


Fig. 16. Approximate representation of dynamic descending hysteresis loop

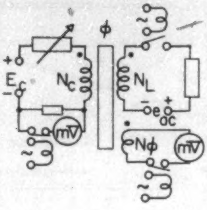


Fig. 17. Measuring circuit for Fig. 15

trol current i_c can be represented approximately as

$$N_c i_c = N_c I_\infty - k\, d\phi/dt \tag{22}$$

and with

$$-d\phi/dt = 2f\Delta\phi \tag{23}$$

the following relation will be obtained:

$$N_c I_c = N_c I_\infty - 2fk\Delta\phi \tag{24}$$

This relation is represented in Fig. 16 by the chain line, which is tangential to the linear portion of the control magnetization curve. The values of I_∞ and k can be obtained from Fig. 16.

For various values of control resistance, the average control current during the resetting half-cycle is measured by the circuit of Fig. 17 and the values of average control current thus obtained agree very closely with the values expected from a consideration of two extreme cases, represented by the current reset corresponding to the control magnetization curve and the direct-voltage reset mentioned previously; see Fig. 15.

Appendix II. Calculations Considering a Parallelogramic Hysteresis Loop

Assume that the core takes the parallelogramic hysteresis loop shown in Fig. 18, then for the descending branch of the hysteresis loop

$$\left.\begin{array}{l} L(i_c - i_0) = N_c(\phi_S - \phi) \\ \text{and } |\phi| < \phi_S \text{ and } d\phi/dt < 0 \end{array}\right\} \tag{25}$$

The equation of control circuit during the control half-cycle for the simplified circuit of Fig. 1(B) is

$$E_c = R_c i_c + L\, di_c/dt \tag{26}$$

Using the conditions

$$i_c|_{t=0} = i_0 \tag{27}$$

$$i_c|_{t=T/2}/2 = i_r \tag{28}$$

the solution of equation 26 yields the following equation

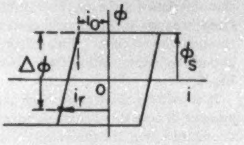


Fig. 18. Parallelogramic hysteresis loop

$$E_c/R_c = (i_r - i_0\epsilon^{-R_c/2fL})(1-\epsilon^{-R_c/2fL})^{-1} \tag{29}$$

From this, the input power of the magnetic amplifier is

$$P_c = R_c(i_r - i_0\epsilon^{-R_c/2fL})^2(1-\epsilon^{-R_c/2fL})^{-2} \tag{30}$$

The control circuit resistance R_{c0}, which minimizes the control power for the specified amount of reset flux, can be calculated as follows:

$$\frac{2R_{c0}}{2fL} = \frac{(i_r - i_0\epsilon^{-R_{c0}/2fL})(1-\epsilon^{-R_{c0}/2fL})}{(i_r - i_0)\epsilon^{-R_{c0}/2fL}} \tag{31}$$

To simplify the calculations, put

$$i_r/i_0 = 1+\delta \tag{32}$$

$$R_{c0}/2fL = \lambda \tag{33}$$

Then equation 31 can be reduced to

$$\delta = \frac{(1-\epsilon^{-\lambda})^2}{2\lambda\epsilon^{-\lambda} - (1-\epsilon^{-\lambda})} \tag{34}$$

When the constants of cores and control winding are known, λ and δ correspond respectively to the optimum value of control circuit resistance and the amount of reset flux. Therefore equation 34 represents the relation between the amount of reset flux $\Delta\phi$ and the optimum circuit resistance.

The minimum control power can be calculated from equations 29 and 30 as

$$P_{c0} = fL\epsilon^{\lambda}(i_r - i_0\epsilon^{-\lambda})^2(1-\epsilon^{-\lambda})^{-1}(i_r - i_0)^{-1} \tag{35}$$

On the other side the reset core loss can be represented as

$$P_r = \frac{fL}{2}(i_r^2 - i_0^2) \tag{36}$$

The ratio of P_{c0} to P_r is

$$\frac{P_{c0}}{P_r} = \frac{2\epsilon^{\lambda}(1+\delta-\epsilon^{-\lambda})^2}{(1-\epsilon^{-\lambda})\delta^2(2+\delta)} \tag{37}$$

The equivalent resistance of the core seen from the control winding during the control half-cycle may be measured as indicated in the circuit of Fig. 5

$$R_{eff}' = E_c/\bar{i}_0 \tag{38}$$

where $\bar{i}_0$ is the average current during the control half-cycle, as

$$\bar{i}_0 = i_0(1+\delta/2) \tag{39}$$

During the resetting interval in the circuit of Fig. 5

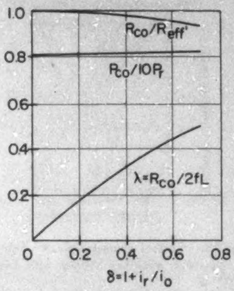


Fig. 19. Calculated results for parallelogramic hysteresis loop

$$E_c = 2fL\, i_0 \times \delta \tag{40}$$

$$\text{and } R_{eff}' = 2fL\delta/(1+\delta/2) \tag{41}$$

Therefore

$$\frac{R_{c0}}{R_{eff}'} = \frac{(\epsilon^{\lambda}-1)(1+\delta-\epsilon^{-\lambda})(2+\delta)}{4\delta^2} \tag{42}$$

Fig. 19 presents the relations of equations 34, 37, and 42. The results indicate very close agreement with the simplified analysis described in the beginning of this paper.

References

1. MAGNETIC AMPLIFIERS (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 269–70.

2. ANALYSIS OF THE IDEAL HALF-WAVE MAGNETIC AMPLIFIER CIRCUIT, T. Kikuchi. *Journal*, Institute of Electrical Engineers, Tokyo, Japan, vol. 75, 1955, pp. 617–21.

3. ON THE MECHANICS OF MAGNETIC AMPLIFIER OPERATION, Robert A. Ramey. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1214–23.

4. SYNTHESIS OF HALF-WAVE MAGNETIC AMPLIFIER CIRCUITS WITH DECREASED INPUT POWER, T. Kikuchi. *Proceedings*, Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., Apr. 1956.

5. THE BIASED RECTIFIER AMPLIFIER–A PULSE MAGNETIC AMPLIFIER, R. E. Morgan. *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 584–90.

6. MEASUREMENT OF CONTROL MAGNETIZATION CURVE OF MAGNETIC AMPLIFIER CORE USING VIBRATING RECTIFIERS, T. Kikuchi, K. Murakami. *Journal*, Institute of Electrical Engineers, Tokyo, Japan, vol. 75, 1955, pp. 746–52.

7. ON THE FLUX CONTROL OF THE SATURATING REACTOR FOR MAGNETIC AMPLIFIERS, T. Kikuchi, K. Murakami. *The Technology Reports*, Tohoku University, Sendai, Japan, vol. 20, pp. 165–95.

8. Storm, *op. cit.*, pp. 58–61.

9. MAGNETIC CHARACTERISTICS PERTINENT TO THE OPERATION OF CORES IN SELF-SATURATING MAGNETIC AMPLIFIERS, R. W. Roberts. *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 682–90.

Rectilinearity of Electron-Beam Focusing Fields from Transverse Component Determinations

P. P. CIOFFI
MEMBER AIEE

IN MAGNETIC focusing of electron beams, the field must be rectilinear and substantially uniform. Rectilinearity is an especially important requirement for traveling-wave tube focusing because small deviations between the magnetic-field and electron-beam axes result in appreciable electron-current interception within the tube. Since the magnetic field cannot be made strictly uniform, it is required to have an axis of symmetry coincident with the mechanical axis of the magnetic structure. The mechanical axis is a convenient reference axis for locating the electron-beam axis. The percentage transmission of electron-beam current through the traveling-wave tube is a measure of the degree of focusing.

The deviation of a magnetic field from a longitudinal axis is determined by the intensity of its perpendicular component to that axis. The origin of transverse components may be caused by magnetic and mechanical asymmetries of the structure, and their evaluation and distribution along the axis provide the means of detecting and correcting their cause. The order of magnitudes involved in such measurements is approximately 1.0 oersted of transverse field in 800 oersteds of longitudinal field, corresponding to an axial deviation of 0.005 inch in 4 inches. These quantities imply an extraordinary degree of precision not only in magnetic-field measurements but also in the location and orientation of the test probe. Small errors of positioning of its magnetic center and alignment of its magnetic axis with respect to the axis of reference result in large errors of measurement because of the component of longitudinal field in the test-probe axis. It can be shown, however, that precise determinations of transverse fields can be made independently of errors of positioning of the test probe by a technique in which these errors cancel out. Furthermore, it will be shown that curves giving the distribution of transverse fields along the axis can be quickly and accurately traced. The interpretation of such curves reveals the mechanical and magnetic asymmetries which give rise to transverse fields, and the effects of applied corrections can be immediately observed.

Definition of Transverse Field

There is no ambiguity in defining the transverse component of a strictly uniform longitudinal field characterized by straight parallel lines of force. If there is an angular displacement between the direction of the magnetic field and a given axis, Fig. 1, the perpendicular component to the given axis is invariable throughout the volume of the field. In this case, correction for the angular deviation requires simply a mechanical adjustment which makes the transverse component vanish at one point.

In practice, however, a degree of field uniformity defined by straight parallel lines of force is not realized. A permanent-magnet focusing structure, Fig. 2, with long air gap and large axial-pole piece holes, has a field-intensity distribution as shown in Fig. 3(A), and the lines of force make curved paths, Fig. 3(B), because of the variation in flux density corresponding to this field-intensity distribution. Since the lines of force are neither straight nor parallel with each other or with the axis, radial components of varying magnitudes are found at all points where they tilt toward the axis. The resultant of these components in any perpendicular plane to the axis is the transverse field at that point. Its distribution may be evaluated from determinations of its components in two perpendicular directions at various points along the axis.

Theory of Transverse-Field Determinations

It is evident from the foregoing discussion that the method of measurement must discriminate between the true transverse field and a component of the longitudinal field caused by errors of positioning of the test probe with respect to the axis of the magnetic structure. The magnitude of this component depends upon the angular deviation of the probe axis from the perpendicular to the field axis. For nonuniform fields, it also depends upon the displacement of its magnetic center from the field axis. Since the elimination of such errors requires a degree of mechanical precision in the construction, positioning, and alignment of the test probe, which is difficult and uncertain of accomplishment, the transverse field cannot be measured directly but must be derived from measurements which nullify these errors. The method used is essentially an exploration of the longitudinal field for axial symmetry.

In Fig. 4, let line XX' represent the axis of a nonmagnetic mandrel which coincides with the mechanical axis of the magnetic structure under test. Assume also that the magnetic axis OA of the test probe, mounted in a transverse opening in the mandrel, makes angle θ with the perpendicular YY', and that its magnetic center is not on the mechanical axis XX'. Since the magnetic field is nonuniform, the upper and lower half of the probe will be in slightly different fields which on the average may be represented by H_θ making angle α with the axis and H_θ' making angle α' with the axis. The resultant field parallel to the probe axis is

$$H = H_\theta \sin(\theta + \alpha) - H_\theta' \sin(\theta - \alpha') \qquad (1)$$

Expressing this field in terms of the longitudinal and vertical components of H_θ and H_θ'

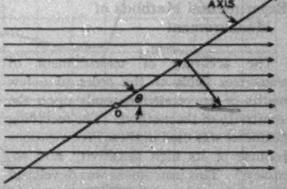


Fig. 1. Transverse field component caused by an angular displacement in uniform longitudinal field

---

Paper 56-518, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Great Lakes District Meeting, Fort Wayne, Ind., April 16–18, 1956, and re-presented at the Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted July 3, 1956; made available for printing December 18, 1956.

P. P. CIOFFI is with Bell Telephone Laboratories, Inc., Murray Hill, N. J.

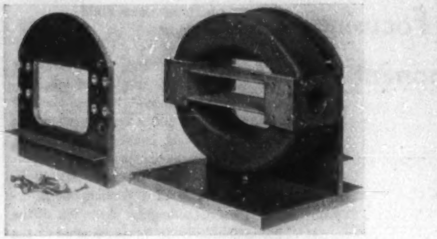


Fig. 2. Permanent magnet for traveling-wave tube electron beam focusing

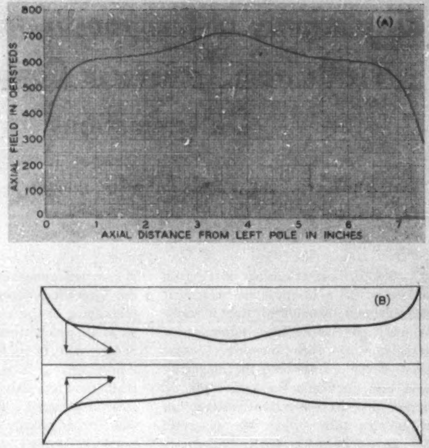


Fig. 3. A—Permanent magnet axial field distribution. B—Transverse components of individual lines of force caused by nonuniform field distribution

$$H = H_l \sin\theta + H_v \cos\theta - H_v' \cos\theta + H_l' \sin\theta \qquad (2)$$

If the mandrel holding the search coil is subsequently rotated 180 degrees about its axis, the probe axis is oriented in the direction OA' and also makes angle θ with the perpendicular YY'. In this new orientation, the resultant field parallel to the probe axis is

$$H' = H_g' \sin(\alpha' + \theta) - H_g \sin(\alpha - \theta) \qquad (3)$$

or

$$H' = H_l \sin\theta - H_v \cos\theta + H_v' \cos\theta + H_l' \sin\theta \qquad (4)$$

From equations 2 and 4

$$(H_v - H_v') \cos\theta = \frac{H - H'}{2} \qquad (5)$$

The transverse field on the axis, by definition, is

$$H_t = H_v - H_v' \qquad (6)$$

and since angle θ is small

$$H_t = \frac{H - H'}{2} \qquad (7)$$

H_t can be evaluated from separate determinations of H and H', or from direct determinations of their difference, $H - H'$. A description of the methods of measurement is given in the following section.

Experimental Methods of Measurement

The accuracy of measurement of transverse fields of the order of a few tenths of an oersted depends upon the sensitivity of the method. The recording fluxmeter was used in these investigations because of its high sensitivity and accuracy.[1] This instrument was developed for tracing magnetization curves and hysteresis loops of magnetic materials. Its integrating circuit, with infinite-time constant, develops an output current proportional to the flux interlinking the search coil connected to it. Hysteresis loops are traced with the X-Y recorder of the instrument as variations of integrator output current with cyclic variations of magnetizing force. In these investigations, the instrument was used to determine the magnetic field distribution in the air gap of magnets. It was used, at first, for point-by-point determinations and, subsequently, for tracing directly longitudinal and transverse field distributions.

POINT-BY-POINT DETERMINATIONS

The probe was a search coil wound on a brass bobbin 0.25 inch in diameter and 0.2 inch long with approximately 2,000 turns of 1.5-mil-diameter Formex-insulated copper wire. It was used as a flip coil to make separate determinations of H and H'. For this purpose it was mounted in a pivoted brass frame which could be flipped through 180 degrees by pulling the appropriate end of a string wrapped around the frame. The search-coil assembly was mounted in a transverse opening in the mandrel. By locating the probe center at various radial distances from the axis of the mandrel, the air-gap volume could be explored over a wide region. The orientation of the coil could also be set for measuring the longitudinal field. The mandrel was supported in accurately fitting bushings, with its axis coincident with the mechanical axis of the magnet structure under test. Its orientation was read on an index head keyed to a slot in the mandrel.

Measurements of H and H' were made by flipping the search coil first with the mandrel in one orientation and then with the mandrel orientation displaced 180 degrees. The determinations were made in the vertical and horizontal direction at various points along the axis.

With the probe fixed transversely in the mandrel, there is no contribution in symmetrical fields from errors of positioning by a 180-degree change in mandrel orientation. The resulting fluxmeter reading, therefore, is a direct measurement of the quantity $H - H'$, and point-by-point measurements can be made by 180-degree flips of the mandrel. This method is simpler and more convenient than the flip-coil method since it requires but one measurement for each determination.

AUTOMATIC TRACING OF TRANSVERSE AND LONGITUDINAL FIELD DISTRIBUTION CURVES

The point-by-point method described in the foregoing is accurate, but it is tedious and laborious because of the many determinations required to evaluate the effectiveness of magnetic and mechanical adjustments on magnetic field rectilinearity. For these reasons, the recording fluxmeter was adapted to trace directly, in a few seconds, curves defining the transverse field distribution over the gap length, automatically and accurately. This permitted the immediate observation of the effect of adjustments.

The method is illustrated in Fig. 5(B). The search coil was fixed in a transverse hole in a 3/8-inch-diameter mandrel. As in the previous case, an index head

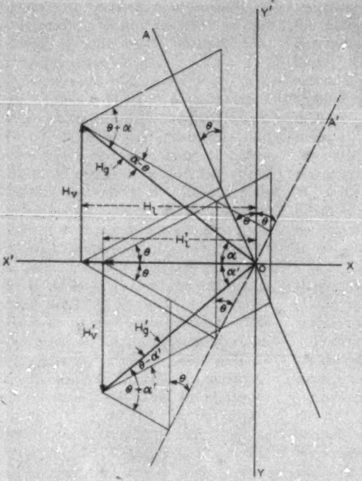


Fig. 4 (left). Analytical representation of the problem of transverse field measurement

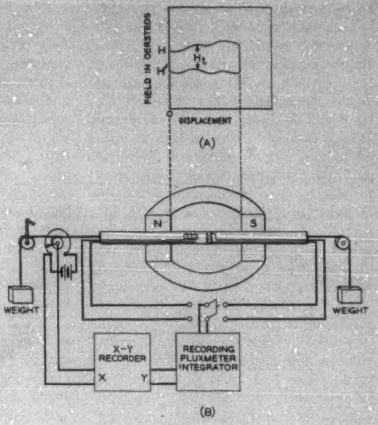


Fig. 5 (right). Method of automatically tracing transverse and axial field distributions

keyed to a slot in the mandrel determined its orientation. The mandrel was accurately mounted in suitable bushings on the mechanical axis of the magnet under study. A fine wire attached to one end of the mandrel was held coaxially with the mandrel by passing it over a pulley and weighting its free end. A second wire was attached to the other end of the mandrel. This wire, also coaxial with the mandrel and with its free end weighted, was wrapped once around the knob of a 10-turn potentiometer and once around a capstan with a hand crank. Turning of the crank produced translatory motion of the rod and a voltage output from the potentiometer proportional to the search-coil displacement. This voltage applied to the *X-Y* recorder of the fluxmeter moved the pen proportionally to the search-coil displacement. The search coil was connected to the integrator of the fluxmeter and its output moved the paper drum proportionally to the field intensity. The recorder thus traced a curve of the variation along the axis of the field intensity *H*, defined by equation 1, as the search coil moved with fixed orientation from pole to pole, Fig. 5(A). If the search coil makes the same traversal with 180-degree change in mandrel orientation, the recorder traces a curve of the variation of the field *H'*, defined by equation 3. The difference in amplitude between the two curves at any point along the axis is a measure of the transverse field at that point. In the absence of transverse field components, the two curves would coincide. (Patterns of transverse field distributions, Figs. 7 and 9, were traced in this manner in approximately 30 seconds.)

The longitudinal field distribution curve of a magnet structure can also be automatically traced with accuracy and rapidity. For this purpose, a second search coil was provided with its axis coincident with the mandrel axis. Longitudinal or transverse field distribution curves were taken by switching the recording fluxmeter to the appropriate coil. In tracing longitudinal field distribution curves, a zero field reference point was established by shielding the search coil.

Results

Transverse field determinations were undertaken for the various electromagnet and permanent-magnet structures under development for electron-beam focusing investigations. Such measurements were necessary for correlating with the degree of focusing, for investigating and correcting the causes of large intensity distributions, and for establishing permissible limits.

Fig. 6 illustrates the transverse field distribution, taken point by point in an early type of solenoidal electromagnet. The relatively high peaks in the distribution curve account for an appreciable amount of electron-beam interception observed in focusing applications of this structure. Electromagnets of the same general type, which were

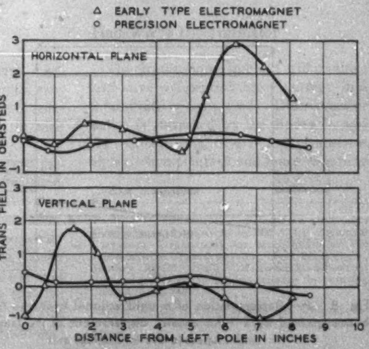


Fig. 6. Comparison of transverse field distributions in two electromagnets

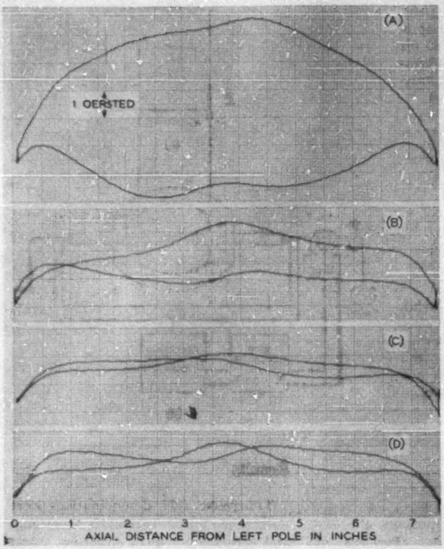


Fig. 7. A—Unidirectional transverse field distribution pattern of a permanent magnet. B—Effect of partial magnetic equalization. C—Horizontal, and D—Vertical plane distribution patterns after adjustments

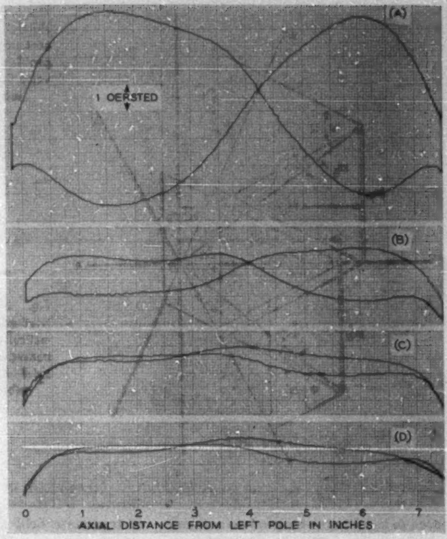


Fig. 9. A—Transverse field distribution with reversal of field direction. B—Effect of partial magnetic equalization. C—Horizontal, and D—Vertical plane distribution patterns after adjustments

designed and constructed for high magnetic and mechanical precision, reflect the effect of these precautions in the exceedingly small transverse fields observed. Maximum values of less than 0.5 oersted were found and, in focusing applications, the electron-beam transmission was greater than 99% (per cent). These field distributions are also illustrated, for comparison, in Fig. 6.

The development of permanent magnets for high-precision focusing of traveling-wave tubes was facilitated by the recognition and correction of the large transverse fields inherent in these structures. Permanent-magnet structures of the type illustrated in Fig. 2, consisting of an upper and lower magnet casting, were designed and constructed to develop average longitudinal fields ranging from 600 to 800 oersteds, with a uniformity of ±5%. Differences in magnetization of the magnet castings of the assembly cause axial deviations of the magnetic field which are observed as strong transverse components in the vertical plane. Differences in their mechanical alignment, on the other hand, cause transverse components principally in the horizontal plane. Examples of transverse field distribution patterns, traced with the recording fluxmeter, for a permanent-magnet structure of this type are shown in Fig. 7. The vertical transverse field, Fig. 7(A), has a unidirectional distribution along the axis with a maximum

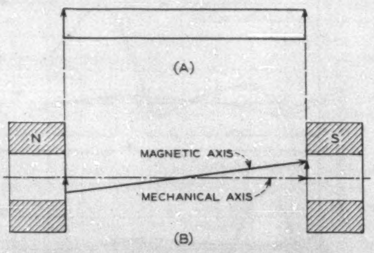


Fig. 8. A—Representation of a unidirectional transverse field distribution, caused by B—An angular displacement of the magnetic axis with respect to the mechanical axis

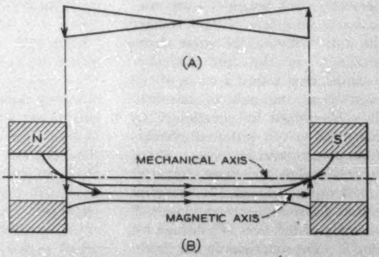


Fig. 10. A—Representation of a transverse field distribution with reversal, caused by B—A lateral displacement of the magnetic axis of a nonuniform field with respect to the mechanical axis

intensity of 6 oersteds near the center of the air gap. Transverse fields of this magnitude cause a large interception of electron current.

A transverse field pattern of this type can be explained by a tilt of the magnetic axis with respect to the mechanical axis, Fig. 8. This angular deviation arises from differences in pole strength between the upper and lower magnet castings of the assembly, which cause the magnetic axis to originate and terminate on opposite sides of the mechanical axis. By equalization of magnetization and adjustment of mechanical alignment, the transverse fields can be reduced to the extent that their effect on electron-beam focusing becomes relatively unimportant. The results of partial equalization are illustrated in Fig. 7(B), and the results of final adjustment for both the horizontal and vertical components are illustrated in Figs. 7(C) and (D). In this adjustment, the maximum intensity of the transverse field distribution is less than 1.0 oersted and makes a number of reversals along the axis. The attainment of such reversals is a desirable feature of the final adjustment since such configurations have a much smaller net effect on electron-beam transmission than unidirectional transverse fields. Electron-beam focusing after these adjustments resulted in approximately 99.95% transmission and illustrates the high degree of longitudinal field rectilinearity that can be achieved.

In another kind of vertical transverse field distribution pattern, Fig. 9, the transverse field reverses in direction near the center of the air gap. This type of distribution may be accounted for by a lateral displacement of the magnetic axis of a nonuniform field from the mechanical axis, as a result of differences in intensity level of magnetization between the two magnet castings of the structure, Fig. 10. The effect of partial magnetic equalization is illustrated in Fig. 9(B), and the result of final adjustments which reduced the vertical and horizontal transverse field components to less than 1.0 oersted is illustrated in Figs. 9(C) and (D). Focusing under these conditions of adjustment resulted also in an electron-beam transmission of approximately 99.95%.

Transverse field distribution patterns are determined by the magnet structure configuration, by the magnetic and mechanical asymmetries of the structure, and by the longitudinal field distribution. A combination of these various effects may result in very complex patterns which do not lend themselves to simple analysis and correction. Corrective measures are therefore required to ensure that transverse field distributions can be held within tolerable limits. A method of transverse field control, applied to the permanent-magnet structures of these investigations, will be discussed in a subsequent publication.

Summary

Transverse field determinations provide a sensitive method of evaluating the degree of rectilinearity of electron-beam focusing fields. The determinations require the measurement of transverse components of a few tenths of an oersted in a total field of 600 to 800 oersteds. An accurate method of measurement has been described which nullifies the error component arising from unavoidable errors of positioning of the test probe with respect to the reference axis. The method is independent of the nature of the probe and is applicable to any method of measuring small fields. Adaptation of the recording fluxmeter to these determinations has made it possible to trace quickly and accurately patterns of transverse field distributions along the axis. The patterns reveal the immediate effects of magnetic and mechanical adjustments on longitudinal field rectilinearity.

The high degree of field rectilinearity obtainable by effective adjustment of transverse field components establishes the permanent magnet as a precision instrument for focusing high-intensity electron beams. This accuracy has made it possible to reduce the field intensity level so that it is closer to the theoretical Brillouin value. Many permanent-magnet structures were produced and accurately adjusted from transverse field pattern determinations. The magnets focused electron beams with greater than 99.95% transmission.

Reference

1. A Recording Fluxmeter of High Accuracy and Sensitivity, P. P. Cioffi. *Review of Scientific Instruments*, New York, N. Y., vol. 21, 1950, pp. 624–28.

Transistorized Phase Discriminators

ALBERT N. DE SAUTELS
ASSOCIATE MEMBER AIEE

Synopsis: This paper discusses half-wave and full-wave phase discriminators which use transistors and diodes to produce phase-sensitive d-c output capable of driving a magnetic amplifier or energizing relays, heating coils, and hydraulic valves. An analysis of the operating characteristics of the transistor in discriminator circuitry is also included, together with a discussion of power considerations and limitations.

IN aeronautical control, where extensive application may be made of phase discriminators with d-c output, use of the transistor in discriminator circuits has definite advantages. The transistorized discriminator would meet minimum size and power requirements, provide amplification with excellent reliability, and perform well over a wide temperature range.

A preliminary study of half-wave germanium transistor phase discriminators displayed adverse sensitivity of this type of discriminator to temperature variation. Subsequent developments in silicon transistors warranted a supplementary study of phase discriminators using this type of transistor. The results indicate that the transistor discriminator can function with excellent stability over a wide temperature range. Power-handling capabilities are extensive and are limited primarily by the maximum voltage and current rating of the transistor.

Half-Wave Discriminator

The circuit shown in Fig. 1 will produce half-wave differential direct current in a 3-terminal or split load. Discriminator action is furnished by the diodes D_1 and D_2. Considering the N P-N transistor shown when points A and B are becoming

---

Paper 57-31, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted September 10, 1956; made available for printing December 12, 1956.

Albert N. De Sautels is with the Minneapolis-Honeywell Regulator Company, Minneapolis, Minn.

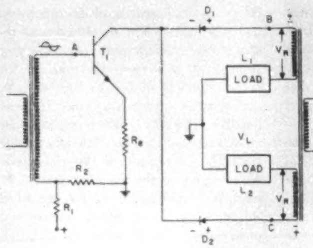


Fig. 1. Half-wave discriminator

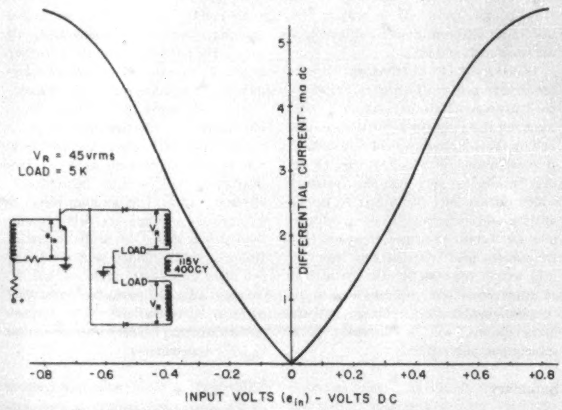


Fig. 2 (below). Response of transistorized half-wave phase discriminator

positive at the same time, T_1 will conduct and current will flow through D_1 and one half of the load L_1. On the negative half-cycle, T_1 is cut off and no current will flow.

If the signal phase at A is reversed and points A and C are becoming positive simultaneously, T_1 will again conduct with current flowing through D_2 and the other half of load L_2. Again, on the negative half-cycle no current will flow.

The resistance R_e may be added to improve transistor interchangeability and performance with temperature variation. The ratio R_1/R_2 would be dependent upon the amplitude of the d-c power supply and the amount of quiescent current acceptable in the load. It is necessary that the diodes D_1 and D_2 be matched for balanced-phase response.

When one diode is conducting, a back voltage appears across the other equal to twice the peak reference voltage minus the voltage drop in the load. Therefore, the nonconducting diode should have an inverse voltage rating in excess of $2\sqrt{2}\,V_R - V_L$.

When the transistor is nonconducting, the collector emitter junction is subjected to an inverse voltage equal to the peak reference voltage. Therefore, the maximum collector voltage rating should be greater than the peak reference voltage value.

Fig. 2 displays representative performance of the half-wave discriminator with a 5-kilohm load. The output power shown, 100 milliwatts, is far from optimum. The relationship of power output to transistor maximum ratings will be discussed later.

Full-Wave Discriminator

There are many instances where it is preferable to use full-wave rather than half-wave discrimination. The large fundamental ripple on the half-wave d-c output would necessitate use of shunting capacitors in hydraulic valve or relay drive applications. Full-wave discrimination would eliminate the need for any capacitor in most instances.

Half-wave magnetic amplifiers present more stable performance with full-wave d-c controls. The greater sensitivity and higher gain of the full-wave magnetic amplifier make its use preferable in many applications. Most efficient control of the full-wave magnetic amplifier can be obtained with full-wave direct current through the control windings.

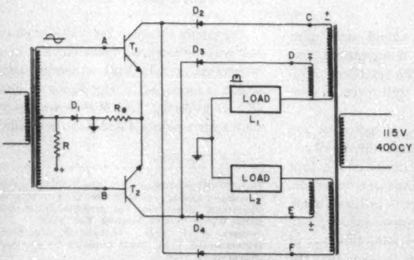


Fig. 3. Full-wave discriminator, 3-terminal load

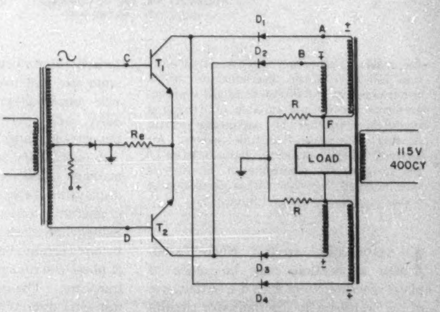


Fig. 4. Full-wave discriminator, 2-terminal load

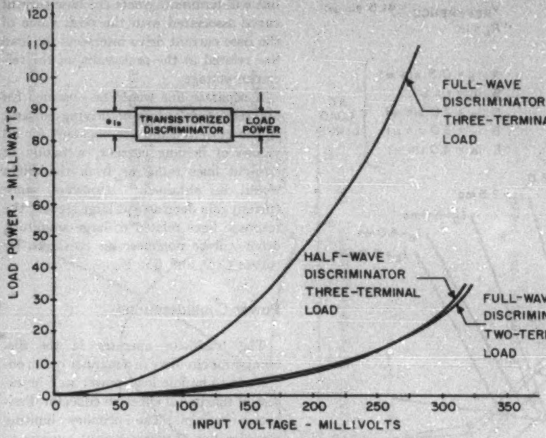


Fig. 5. Load power as function of input voltage

Transistors may be used as well in full-wave discriminator circuitry as in half-wave circuitry.

Three-Terminal Load

An adaptation of the previously discussed half-wave discriminator to full-wave output with a 3-terminal load is shown in Fig. 3. Full-wave output is obtained through the summation of two half-wave actions.

When *A* and *C* become positive simultaneously, T_1 conducts through D_2 and load L_1. During this half-cycle, T_2 is cut off because of the negative signal at *B*. On the second half of the cycle *A* and *C* become negative, cutting off T_1, while *B* and *D* become positive. T_2 then conducts through D_3 and load L_1. Thus, the load experiences two half-wave current pulses during the cycle, or a full wave of current. For a 180-degree phase change in the signal, *A* and *F* would become positive simultaneously, and conduction would occur through T_1, D_5, and load L_2 for one half-cycle. For the second half-cycle *A* and *F* would become negative while *B* and *E* become positive, and T_2 would conduct through D_4 and L_2, completing the full wave of current.

Since both transistors conduct during the cycle to provide an average or effective current, the transistors need not be matched. The same inverse voltage considerations associated with the half-wave circuitry apply to full-wave circuitry. For balanced-phase response, the diodes may have to be matched. Excellent response has been obtained, however, with the production run of unmatched diodes.

Diode D_1 is used to provide a constant bias voltage for the transistors. Control of forward current through D_1 establishes a desired transistor bias voltage related to the forward diode voltage drop. The desired bias voltage is selected in accordance with the amplitude of quiescent load current permissible for a particular application.

An additional function of D_1 is to provide some compensation for temperature changes. Transistor characteristics are such that, with temperature decrease, an increase in bias voltage is required to maintain constant collector current conditions. Conversely, with temperature increase, a decrease in bias voltage is necessary to maintain constant collector current. The temperature characteristics of the bias diode are such that the forward resistance increases with decreased temperature and decreases with increased temperature. Therefore, with a constant current source to the diode, the forward voltage across the bias diode D_1 would increase with lower ambient and decrease with higher ambient temperature. In this way, the drop across the diode would tend to compensate for the changing bias requirements of the transistor. The resistance R_3 may be added to improve transistor interchangeability.

Two-Terminal Load

It may sometimes be desirable to use 2-terminal phase-sensitive d-c output from the discriminator. This can be accomplished with the full-wave discriminator, just described, by providing bleeder resistors and connecting the load as shown in Fig. 4.

For the best power transfer to the load, *R* is approximately equal to the load resistance. One-sixth of the output power is dissipated in the load; the remainder is lost in the bleeder resistors. Thus, this configuration is considerably less efficient than the 3-terminal usage.

The current through the load will be full wave with direction related to signal phase. When point *C* becomes positive, T_1 conducts through D_1; the current takes two paths, one through the load and the other through *R*. On the next half-cycle, points *D* and *B* become positive, and T_2 conducts through D_2 and completes the full wave of current through the load. With reversed signal phase, conduction would be through D_4 and D_3 with a reversal of current direction through the load.

Fig. 5 shows comparable load power for half- and full-wave 3-terminal and full-wave 2-terminal transistor phase discriminators as a function of input voltage. These curves indicate a favorable comparison between half-wave 3-terminal output and full-wave 2-terminal output.

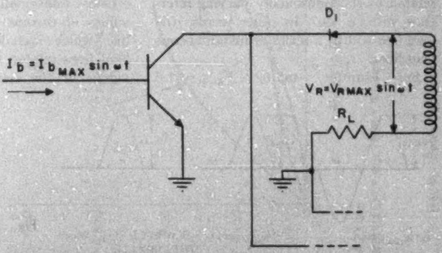


Fig. 6. Part of the phase-discriminating circuit

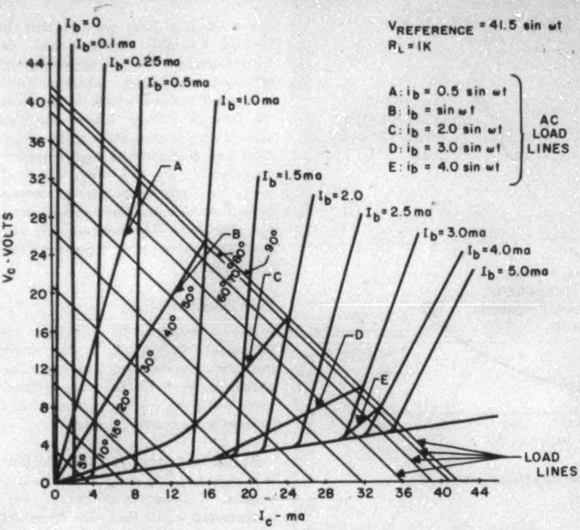


Fig. 7. Characteristics of a grounded emitter transistor in a half-wave discriminator action

Characteristics of the Transistors

The transistor, when used in phase discriminator processes, displays some of the characteristics of both switching and class *B* operation. In the discriminator action, the *NPN* transistor conducts when both the base and the collector are positive. This condition occurs during one half-cycle. During the negative half-cycle the transistor is cut off. Fig. 6 shows a part of the phase-discriminating circuit.

The reference voltage, which provides the collector voltage, varies sinusoidally and is equal to $V_{R\max} \sin \omega t$. For a given load, the load line superimposed on the transistor common-emitter collector characteristics of Fig. 7 will be a series of lines with constant slope; however, perpendicular distance to the origin will be related to the sinusoidally varying reference voltage V_R. In other words, the load line will be a series of instantaneous load lines.

For example, taking $V_{R\max}=41.5$ volts and $R_L=1.0$ kilohm (neglecting, for convenience, the diode drop), 30 degrees after start of the cycle, $V_R=20.7$ volts, and the load line intersects the x and y axes at 20.7 ma (milliamperes) and 20.7 volts respectively; 60 degrees after start of the cycle, $V_R=35.9$ volts, and the load line intersects the x and y axes at 35.9 ma and 35.9 volts respectively.

The input current I_b is sinusoidal and equals $I_{b\max} \sin \omega t$. This indicates that, at the start of the transistor's conducting half-cycle, $V_R=I_b=I_c=V_c=0$. Fig. 8 displays wave forms relating, on a time axis, the input current I_b to the collector voltage and current.

Referring again to Fig. 7, $I_b=V_c=I_c=0$ at the start of the cycle; therefore, the start of the a-c load line will be at the origin. As the driving current I_b increases sinusoidally, the various time values of current will progress along a line which includes intersections with load lines related to corresponding time values of the reference voltage. This line will terminate where the base current curve associated with the peak value of the base current drive intersects the load line related to the peak value of the reference voltage.

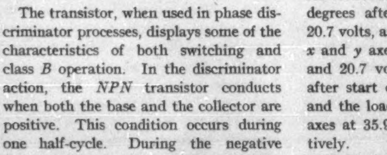

$I_{b_{MAX}}$
π
π
$V_{C_{MAX}}$
π
π
$I_{C_{MAX}}$
π
π
$I_b=I_{b_{MAX}} \sin \omega t$
$V_C=V_{R_{MAX}} \sin \omega t - I_C R_L$ WHERE $I_C \cong I_{C_{MAX}} \sin \omega t$

Fig. 8. Discriminator wave forms

A separate line would be obtained for different peak values of driving current. If the current gain was constant for all values of driving current, a family of straight lines radiating from the origin would be obtained. However, since current gain decreases at large signals, the response lines related to large-amplitude drive will be nonlinear, as so shown by curves *C*, *D*, and *E* in Fig. 7.

Power Considerations

The transistor operates in the discriminator circuitry in a manner which enables it to handle load power well in excess of maximum effective collector dissipation ratings. The primary limiting consideration of power output appears to be the maximum current and the voltage rating of the transistor in so far as these ratings contribute to the maximum collector dissipation rating.

As an example of effective power dissipation per transistor, take a value of reference voltage $V_R=30$ volts rms and $R_L=1.0$ kilohm. The transistor conducts for one half of the cycle and is cut off for the other half-cycle. For an effective differential or load current of 6 ma for the full-wave discriminator, the current furnished by each transistor during its conducting half-cycle would be 6 ma effective or a peak current of 8.48 ma. During the conducting half-cycle, therefore, there would be dissipated in the transistor an effective power equal to the product of the effective current and the rms collector voltage V_c, where $V_c=V_{ref}$ rms$-V_{\text{diode}}-I_{L\text{eff}}\ R_L$. Considering $V_{\text{diode}}=1.0$ volt for this example, then $V=30-1-(0.006)\ (1000)=23$ volts rms; and the effective transistor power dissipated during the conducting half-cycle is $P_{T\text{eff}}=(0.006)\ (23)=138$ milliwatts. Averaging over a full cycle produces for the transistor an effective dissipation of 69 milliwatts. The peak power dissipated in the transistor would equal $\sqrt{2}\,I_{\text{eff}}\times\sqrt{2}\,V_c=276$ milliwatts.

Thus, the following relationship between effective power dissipated in the transistor over one full cycle and the effective load current may be stated for the full-wave discriminator.

$$P_T=\frac{I_L(V_R-V_D-I_LR_L)}{2}=\frac{P_{\text{pk}}}{4}$$

$$=\frac{1}{2}\left(\frac{I_{L_{\text{pk}}}}{\sqrt{2}}\right)\left(\frac{V_{c_{\text{pk}}}}{\sqrt{2}}\right) \quad (1)$$

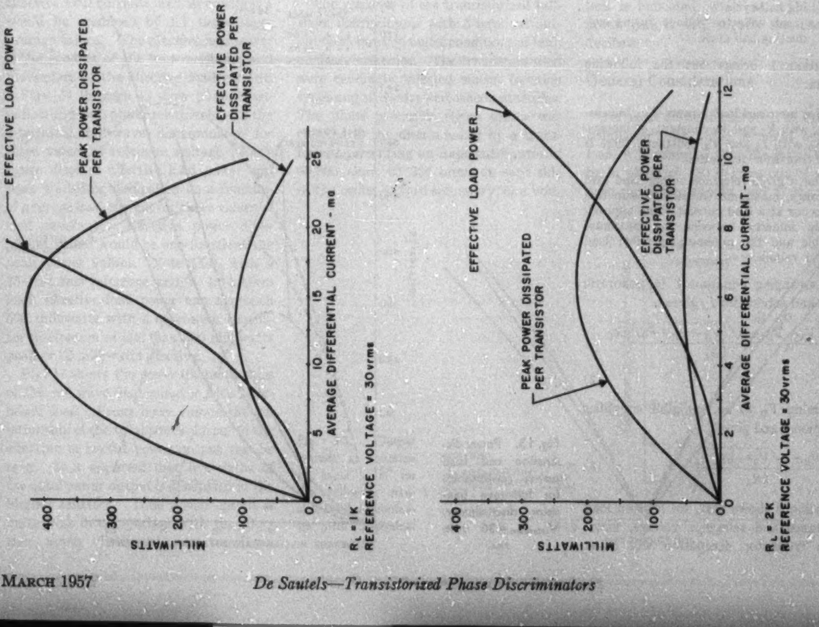


Fig. 9. Power versus current relationships for full-wave transistor discriminator

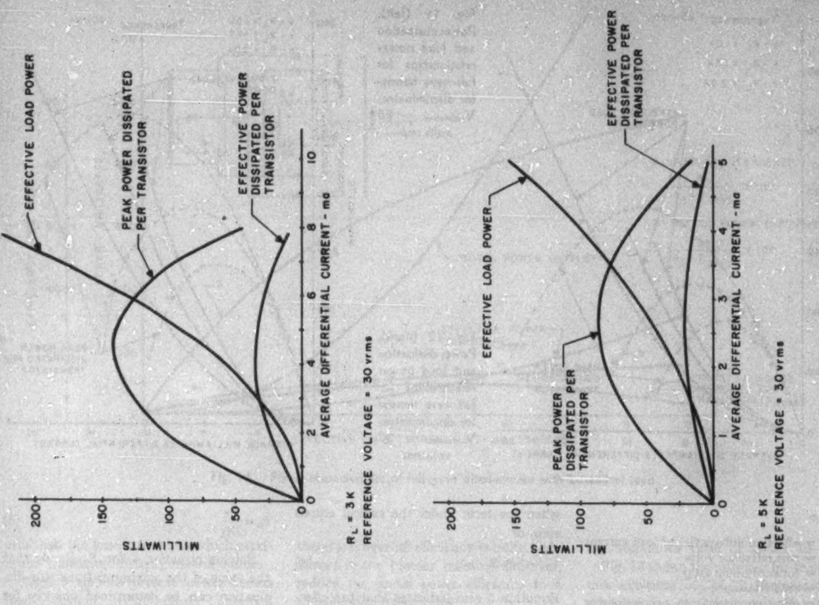


Fig. 10. Power versus current relationships for full-wave discriminator

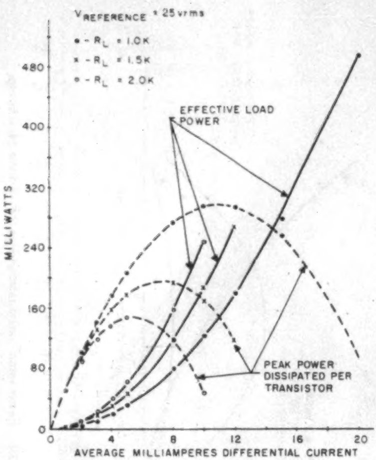
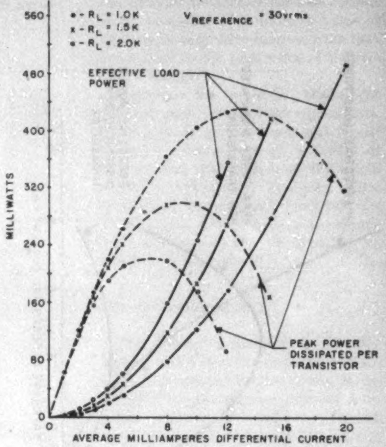


Fig. 11 (left). Power dissipation and load power relationships for full-wave transistor discriminator, $V_{reference}$ = 25 volts rms

Fig. 12 (right). Power dissipation and load power relationships for full-wave transistor discriminator, $V_{reference}$ = 30 volts rms



where

I_L = effective differential or load current
V_R = rms reference volts
V_D = forward diode drop
R_L = load resistance
P_T = effective dissipation per transistor averaged over full-cycle
P_{pk} = peak power dissipated per transistor during conducting half-cycle
I_{Lpk} = peak load current during conducting half-cycle
V_{cpk} = peak collector voltage during conducting half-cycle

Equation 1 brings out the following points:

1. For increased load powers (i.e., increasing I_L with R_L and V_R constant, or increasing R_L with I_L and V_R constant), there is lower transistor dissipation.
2. For a given reference voltage and load resistance, maximum transistor dissipation will occur at a load current value approximately midway between the maximum possible and the minimum possible load current values.

Rearranging equation 1 in quadratic form and solving for I_L gives

$$I_L = \frac{(V_R - V_D) \pm \sqrt{(V_R - V_D)^2 - 8P_T R_L}}{2R_L} \tag{2}$$

Assuming V_D to be negligible simplifies equation 2 and gives

$$I_L = \frac{V_R \pm \sqrt{V_R^2 - 8P_T R_L}}{2R_L} \tag{3}$$

Equation 3 shows that, for a given load resistance and reference voltage, maximum transistor dissipation will occur when the term under the radical equals zero, or

$$V_R^2 = 8P_T R_L \tag{4}$$

Equation 3 also indicates that the effective load current related to this maximum dissipation point is

$$I_L = \frac{V_R}{2R_L} \tag{5}$$

Solving equation 4 for P_T shows that the value of the maximum transistor dissipation can be determined quickly for any combination of values for V_R and R_L. Utilization of equation 5 will then

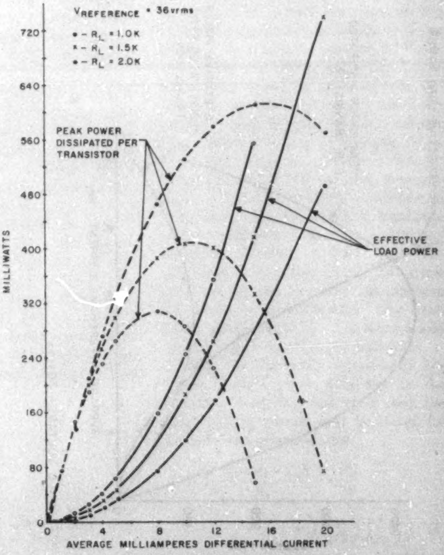


Fig. 13. Power dissipation and load power relationships for full-wave transistor discriminator, $V_{reference}$ = 36 volts rms

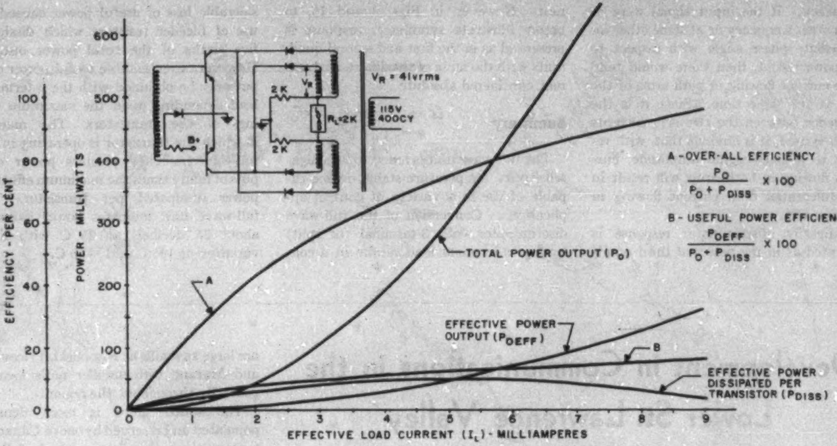


Fig. 14. Power characteristics of full-wave discriminator with 2-terminal load

establish the load current at which maximum transistor dissipation will occur.

Figs. 9 and 10 present power versus load current relationships obtained through use of equation 1 for various load resistance values. The load current axis is calibrated in average values. The effective load currents used in equation 1 would be produced by 1.1 times these average values. The effective load power is the product of the load resistance and the square of the effective load current.

Figs. 11 through 13 show power dissipation and load power relationships of the 3-terminal full-wave discriminator for three values of reference voltage. Each figure displays effective load power and peak transistor dissipation as a function of average load current for three values of load resistance. Effective power dissipation values would be one fourth of the peak power values. Note that, with a 25-volt rms reference and a 1.0-kilohm load, effective load power can approach 500 milliwatts with a maximum transistor dissipation of less than 300 milliwatts peak or 75 milliwatts effective.

Fig. 14 shows the power characteristics of the full-wave discriminator with 2-terminal load. From these curves the relationship of the total power output to the effective or useful power output can be seen. It is apparent that five sixths of the total power output is dissipated in the bleeder resistors. Total power output is quite high in comparison with the effective power dissipated per transistor, therefore, over-all efficiency is quite high. Losses in the bleeder resistors, however, reduce the useful power efficiency to a considerably lower value.

Temperature Characteristics

The response of the transistorized full-wave discriminator with 3-terminal output was observed under conditions of temperature extremes. The transistors used were randomly selected silicon junction types and all diodes were silicon junctions. The phase reversible signal source was coupled to the discriminator by a transformer presenting an impedance ratio of 50,000 ohms to 330 ohms on each side of the center-tapped secondary, or a voltage step-down ratio of about 12.

Fig. 15 shows the response for temperature extremes of 125 C (degrees centigrade) and −55 C as compared with 25 C. Note that considerable power gain is present in this circuit. At 25 C and 125 C a power gain of more than 25 decibels is indicated, while at −55 C the power gain is only slightly less than 25 decibels.

General Considerations

For most efficient operation of the discriminator with three terminal or split load, it is essential that the input signal be in phase or 180 degrees out of phase with the power supply, and at the same

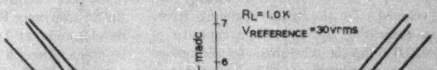


Fig. 15. Output current as function of input volts for the full-wave transistorized discriminator with 3-terminal output

frequency. If the input signal were at some other frequency or at some other intermediate phase angle with respect to the power input, then there would tend to be current flowing in both arms of the load at the same time. Since it is the difference between the two arm currents which is used, it is obvious that, with respect to a given signal amplitude, currents flowing in both arms will result in less differential than current flowing in just one arm.

Ordinarily, discriminator response is presented as in the first and third quadrant. However, in Figs. 2 and 15, to better illustrate symmetry, response is presented as in the first and second quadrants with the sense of the differential current considered absolute.

Summary

The full-wave unit is found to be a high-sensitivity temperature-stable device capable of use in a variety of control applications. Conversion of the full-wave discriminator with 3-terminal (or split) load to 2-terminal load results in a considerable loss of useful power caused by use of bleeder resistors which dissipate five sixths of the total power output. However, a respectable useful power output may be obtained with the 2-terminal load depending upon the maximum ratings of the transistors. The manner in which the transistor is operating in the full-wave circuitry permits power outputs of many times the maximum effective power dissipated per transistor. The full-wave unit indicates power gains of about 25 decibels at 25 C with little variation at 125 C and −55 C.

Development in Communications in the Lower St. Lawrence Valley

F. C. DOAK
ASSOCIATE MEMBER AIEE

J. R. TENNET
NONMEMBER AIEE

THIS PAPER describes the communication facilities in the Lower St. Lawrence region of the Province of Quebec, Canada. An outline is given of the geography of the region which the communication system serves.

Geography

The area under consideration, shown in Fig. 1, comprises both shores of the lower St. Lawrence River. The watersheds on the south shore are steep and quite short as compared with the long rivers on the north shore. For example, the drainage basin of the Manicouagan River is more than 300 miles long. This vast area provides this river with a potential hydroelectric power of 2,000,000 hp (horsepower) 77,000 of which is already developed. On the nearby Bersimis River a generating station with a capacity of 450,000 hp is scheduled to commence operation next year. The water power of the other large rivers in the area is shown on the map. The utilization of these hydraulic resources is a significant factor in the industrial development of the region.

At the extreme north of the region there are practically unlimited deposits of high-grade iron ore at Knob Lake. A railroad 360 miles in length was especially constructed to transport this ore to the port at Seven Islands from where it is shipped by boat to blast furnaces located in many countries throughout the world. In 1955 7,000,000 tons of iron ore were shipped on this railway, the longest railroad built in North America during the past 40 years. Production is expected to be stepped up to 12,000,000 tons in 1956. At present, more than 600 railroad cars, each filled with 80 tons of ore, leave Knob Lake for the port at Seven Islands every day during the ore-shipping season which lasts from the beginning of May to the middle of November.

Further east along the north shore of the St. Lawrence River there is a titanium development at Lac Allard. A railroad 30 miles in length has been built to bring out the ore to the port at Havre St. Pierre.

There are extensive pulpwood operations on both shores of the St. Lawrence. The rate of "cut" is controlled by the Provincial Government, with a view to keeping the industry on a perpetual basis. Part of this pulpwood is manufactured into paper at mills located at Baie Comeau and Chandler. There is a ground pulpwood mill at Clarke City. The remaining pulpwood is shipped to mills outside the region in Canada, the United States, and Europe. Part of the timber on the south shore is processed into lumber. There are large sawmills at Rimouski, Priceville, and Matane with smaller mills located generally throughout the region.

The south shore is more densely populated and is served by one of Canada's transcontinental railways which follows the river edge as far east as Mont Joli, thence cutting across the neck of the peninsula, and again following the shore of the Bay of Chaleur to the town of Gaspé located on a narrow bay at the extreme end of the peninsula. Besides the forestry mentioned heretofore, other industries in this area include farming, manufacturing, fishing, and mining. A copper mine at Murdochville, located in the center of the peninsula, recently commenced production with an output of 125 tons of copper a day.

During the past 20 years the resources of the region have been developed at an ever-increasing rate and, consequently, it has been necessary to extend communication facilities to keep apace with this industrial development. Distances are, of course, a major factor in providing communication service. The land line which skirts the north shore of the St. Lawrence between Tadoussac and Red Bay in Labrador has a pole-line mileage of 877 miles. This gives an idea of the circuit mileages in this region.

The communication facilities in the region are discussed under the following headings:

The communication companies serving the region.
Co-ordination with direct distance dialing plan.
Toll centers.
Wire lines.
Radio telephone.
Carrier telephone equipment.
Dial exchanges.
Buildings.
Future plans.

Paper 57-10, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 24, 1956; made available for printing November 1, 1956.

F. C. DOAK and J. R. TENNET are with the Québec Téléphone Company, Rimouski, Que., Canada.

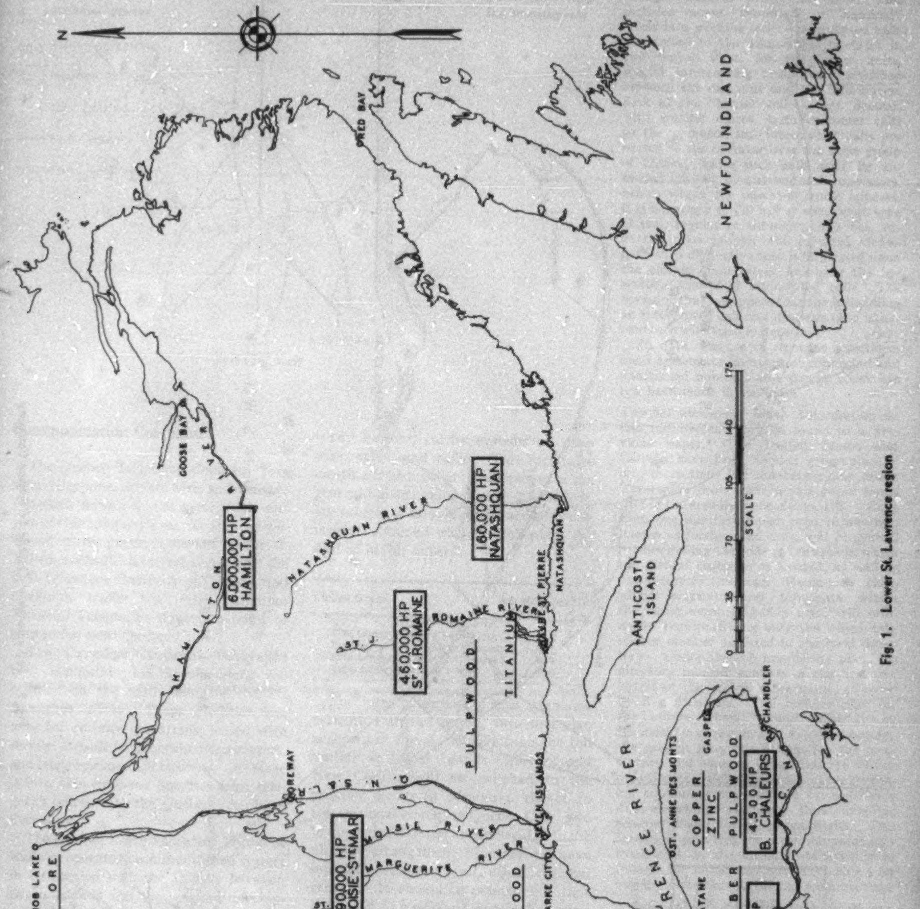

N
KNOB LAKE
IRON ORE
HAMILTON RIVER
GOOSE BAY
6,000,000 HP HAMILTON
RED BAY
NEWFOUNDLAND
NATASHQUAN RIVER
160,000 HP NATASHQUAN
NATASHQUAN
HAVRE ST. PIERRE
ROMAINE RIVER
ST. J.
460,000 HP ST. J. ROMAINE
PULPWOOD
TITANIUM
Q.N.S.&L.R.
OREWAY
SEVEN ISLANDS
MOISIE RIVER
MARGUERITE RIVER
ST.
90,000 HP MOISIE-STEMAR
PULPWOOD
CLARKE CITY
2,000,000 HP MANICOUAGAN
MANICOUAGAN RIVER
PULPWOOD
BAIE COMEAU
120,000 HP OUTARDES
OUTARDES RIVER
2,000,000 HP BERSIMIS
BERSIMIS RIVER
FORESTVILLE
PULPWOOD
2,320,000 HP SAGUENAY
ST. LAWRENCE RIVER
MATANE
STE. ANNE DES MONTS
COPPER
ZINC
PULPWOOD
LUMBER
RIMOUSKI
36,000 HP MITIS
TROIS PISTOLES
45,000 HP RIMOUSKI
I.R.N.C.
GASPE
45,000 HP B. CHALEURS
CHANDLER
C.N.R.
ANTICOSTI ISLAND
0
35
70
105
140
175
SCALE

Fig. 1. Lower St. Lawrence region

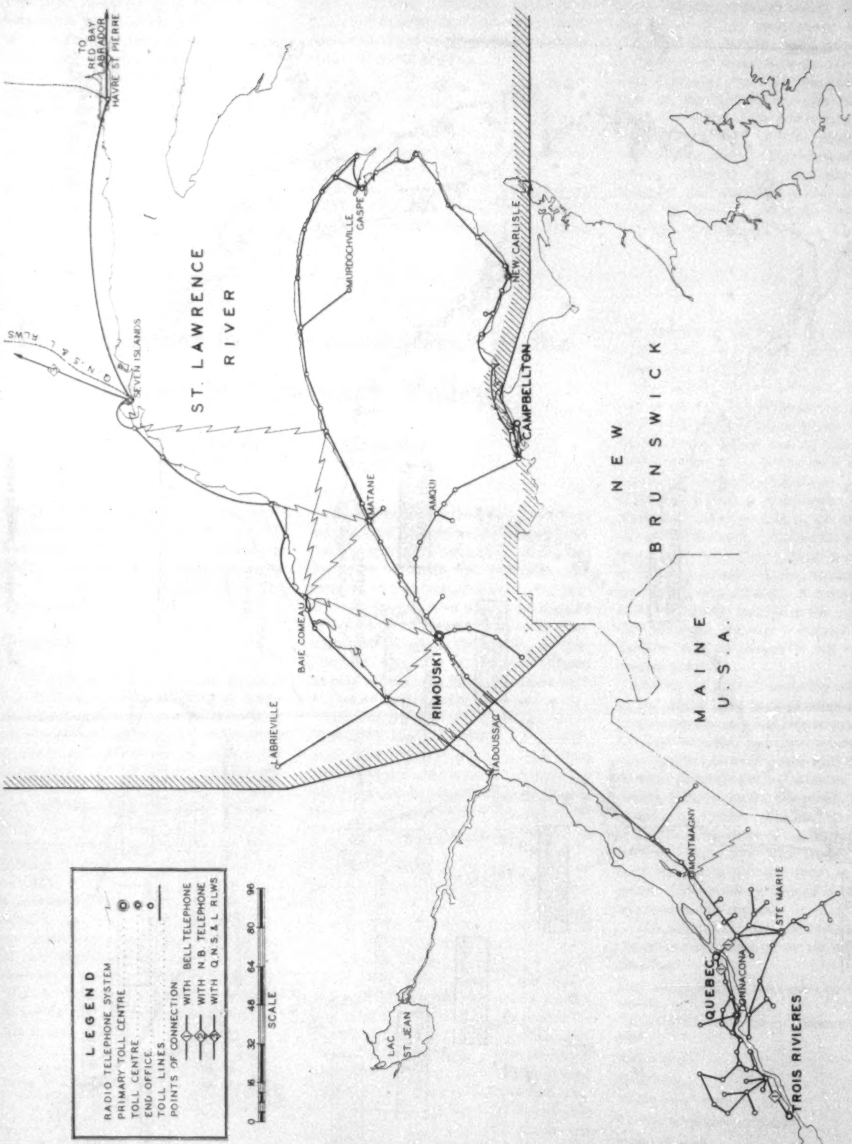

LEGEND
RADIO TELEPHONE SYSTEM
PRIMARY TOLL CENTRE
TOLL CENTRE
END OFFICE
TOLL LINES
POINTS OF CONNECTION
WITH BELL TELEPHONE
WITH N.B. TELEPHONE
WITH Q.N.S.&L RLWS
SCALE
0
16
32
48
64
80
96
ST. LAWRENCE
RIVER
NEW
BRUNSWICK
MAINE
U.S.A.
TO
RED BAY
LABRADOR
HAVRE ST PIERRE
SEVEN ISLANDS
Q.N.S.&L RLWS
MURDOCHVILLE
GASPE
NEW CARLISLE
CAMPBELLTON
MATANE
AMQUI
BAIE COMEAU
RIMOUSKI
LABRIEVILLE
TADOUSSAC
LAC
ST. JEAN
MONTMAGNY
STE MARIE
QUEBEC
DONNACONA
TROIS RIVIERES

Fig. 2. Toll network of Québec Téléphone Company

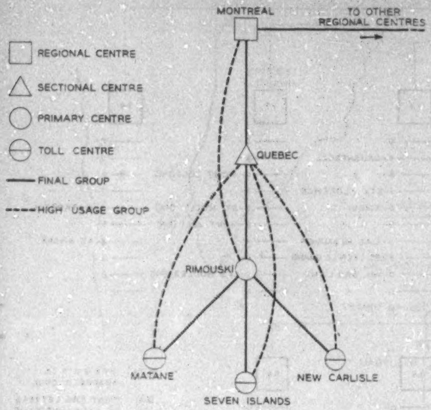


Fig. 3. Montreal regional center homing arrangements

Communication Companies

The Québec Téléphone Company provides telephone, private wire, and message telegraph service in the area. The section of this company's service area, as discussed in this paper, is marked on Fig. 2. Telephone traffic is interchanged with the Bell Telephone Company of Canada, and telegraph traffic with the Canadian National Telegraphs through a point of connection near Quebec City.

The Canadian National Telegraphs have communication circuits along the railroad on the south shore of the St. Lawrence River. These facilities are used for railway operations, leased wire service, broadcast program transmission, and telegraph message business. At some points, this company operates joint telegraph offices with the Québec Téléphone Company.

The Quebec North Shore and Labrador Railway operates a communication system in connection with the railway between Seven Islands and the iron-ore development at Knob Lake. This system is used for railway operations and public telephone and telegraph service to points along the railway. These facilities were described in a previous paper.[1]

Co-ordination with Direct-Distance Dialing Plan

The telephone system in the area is being modernized and extended in accordance with the technical requirements needed for participation in the continent-wide direct-distance dialing plan. Intertoll trunks are mainly of the high-velocity type; namely, carrier systems on open wire, cable, and radio. This facilitates meeting transmission requirements of the general toll-switching plan.[2] The following table is a sample of the actual transmission obtained with the toll plant described in this paper.

Circuit Group	Via Net Loss, Db
Toll center-Primary center	3.0
Primary center-Sectional center	3.7
Primary center-Regional center	3.4
Toll-center-Sectional center	3.0

The interconnection of this telephone network with the general switching plan is shown in Fig. 3. Traffic from the toll centers at Seven Islands, Matane, and New Carlisle will be switched to the primary center at Rimouski, thence to the sectional center at Quebec City, and to the regional center at Montreal. In the ultimate arrangement, Montreal will have intertoll trunks to all other regional centers. It should be pointed out that Rimouski, as a primary center, also serves as the toll center for a number of tributary offices to which it has direct connections.

Constant progress is being made in the conversion of local offices from manual to dial service. To meet the requisites of continent-wide dialing, each local exchange, on its conversion to dial operation, has the following features.

Terminal-per-station equipment. The alternative arrangement, terminal per line, gives an unsatisfactory grade of intercepting service and is very wasteful of central office codes.

Intercepting. Provision is made for the interception by an operator at the toll center of calls to every vacant, disconnected, or changed number within an office. This includes vacant "thousand" or "hundred" levels in a working unit. Intercepted calls are routed to the information operator in two ways. First, for the larger units, special intercepting trunks are provided between the dial unit and the information desk at the "home" toll center. Second, with smaller offices, both customer calls to the operator and intercepted calls are routed to the operator over the same group of trunks. Since such calls may be answered initially by any one of the operators, before whom the operator trunk appears, it is necessary to tell her at once what type of call, regular or intercept, she has answered. To provide the required identification, a distinctive tone is impressed upon the circuit immediately following the operator's answer of intercepted calls. The operator thus recognizes that she is handling an intercepted call and disposes of it in accordance with local instructions.

To date, the use of recorded announcement systems to intercept disconnected and unassigned numbers and vacant codes has not been made in the area.

Two-five numbering plan. Information on this requirement may be found in a previous paper.[3] The United States and Canada have been divided geographically into more than 100 numbering-plan areas. The region dealt with in this paper forms part of the area having code no. 418. Calls from one numbering-plan area to another, known as foreign-area calls, will in general require dialing the code of the area in which the distant customer is located, as well as his directory number. Home-area calls, which originate and terminate within the home numbering-plan area, will ordinarily require dialing only the called subscriber number as listed in the local directory. Under the 2-5 numbering plan, the directory number consists of the first two letters of the central office name, followed by five numerals. The number plan for the offices which "home" directly on Rimouski is shown on Fig. 4. For example, the present directory listings for the community dial office at Mont Joli are PResident 5-xxxx. Substantial savings in switching equipment are obtained by employing digit-absorbing switches, thus avoiding selector stages for all seven digits. This feature is provided for, at a comparatively small cost, in the initial installation even though the office may not be scheduled for connection to the general plan for some time.

Signals required for continent-wide dialing. All switching equipment in the region is arranged to return busy flashes and tones. These signals are well established by general usage.

Station equipment. All new dial station equipment is being equipped with 2-5 number cards and dial number plates with letters and numerals.

MANUAL TOLL SWITCHBOARDS

In the conversion to dial program, no. 3C switchboard equipment is installed at the toll centers. The positions are supplied, wired for key pulsing, and when the number of sections reaches 15 the economic aspects of key pulsing for the

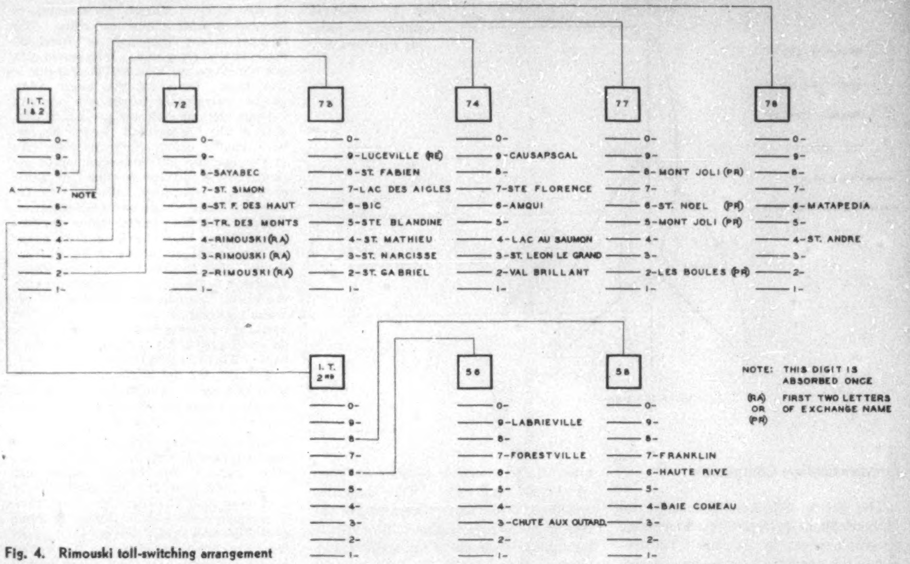

I. T. 1 & 2
0-
9-
8-
A 7- NOTE
6-
5-
4-
3-
2-
1-
72
0-
9-
8-SAYABEC
7-ST. SIMON
6-ST. F. DES HAUT
5-TR. DES MONTS
4-RIMOUSKI (RA)
3-RIMOUSKI (RA)
2-RIMOUSKI (RA)
1-
73
0-
9-LUCEVILLE (RE)
8-ST. FABIEN
7-LAC DES AIGLES
6-BIC
5-STE BLANDINE
4-ST. MATHIEU
3-ST. NARCISSE
2-ST. GABRIEL
1-
74
0-
9-CAUSAPSCAL
8-
7-STE FLORENCE
6-AMQUI
5-
4-LAC AU SAUMON
3-ST. LEON LE GRAND
2-VAL BRILLANT
1-
77
0-
9-
8-MONT JOLI (PR)
7-
6-ST. NOEL (PR)
5-MONT JOLI (PR)
4-
3-
2-LES BOULES (PR)
1-
78
0-
9-
8-
7-
6-MATAPEDIA
5-
4-ST. ANDRE
3-
2-
1-
I. T. 2nd
0-
9-
8-
7-
6-
5-
4-
3-
2-
1-
56
0-
9-LABRIEVILLE
8-
7-FORESTVILLE
6-
5-
4-
3-CHUTE AUX OUTARD
2-
1-
58
0-
9-
8-
7-FRANKLIN
6-HAUTE RIVE
5-
4-BAIE COMEAU
3-
2-
1-
NOTE: THIS DIGIT IS ABSORBED ONCE
(RA) OR (PR) FIRST TWO LETTERS OF EXCHANGE NAME

Fig. 4. Rimouski toll-switching arrangement

particular office is studied. An information desk is at the end of the switchboard line-up. Slack cable to this desk is coiled in the 3*C* sections, permitting the information desk to be conveniently moved along as the switchboard grows. Other miscellaneous features include speed of answering-time recorders and in-board operator-training equipment.

Toll Centers

A toll center is defined as a long-distance switching center serving directly a group of local offices. Fig. 5 shows the four toll centers in the region; namely, Rimouski, Matane, Seven Islands, and New Carlisle. As mentioned previously, Rimouski is also the primary center for the area. Each of these centers is surrounded by a cluster of community exchanges. In the final plan, each of these communities will be served by an unattended community dial office (CDO), with direct trunks to its home toll center. The exchanges which have already been converted to dial operation are designated as such in Fig. 5. Under this arrangement any exchange in a toll center area can reach any other office in the same area by only one switch at the toll center. This provides faster service and better transmission than the older arrangement whereby it was necessary to pass through several operators to span a comparatively short distance. Moreover, there is a further benefit, i.e., a higher utilization of facilities is obtained by concentrating all traffic to and from an exchange over a single group of circuits. High-usage trunks are provided between the toll centers and thus any point in the region will be reached with a maximum of two switches.

Traffic to and from the region is routed through the primary center at Rimouski to the sectional center at Quebec City and thence to the comprehensive toll network which serves the United States and Canada.

These toll center areas, like others throughout the continent, are the basis upon which the general toll-switching plan is built.

Wire Lines

LOCAL LINES

For local service from the telephone exchange to the subscriber's premises the following materials are used: paper-insulated lead-covered cable, type *B* rural distribution cable (six plastic-insulated pairs surrounding a steel messenger wire), high-strength galvanized-steel wire, and neoprene-jacketed parallel-drop wire.

Since 1953 new construction of subscriber lines has been planned in accordance with the new resistance design. This uses ohms rather than decibels for cable-gauge selection. Substantial savings and improved transmission have resulted from the use of these new practices, e.g.,

1. Use of minimum copper in subscriber's loop plant.
2. Maximum utilization of the improved transmission and dialing characteristics of the new telephone instruments such as the *500*-type set.
3. Conservation of engineering time.

To preclude the use of more than one gauge of conductor in the same cable and to keep the resistance within requirements, a subscriber's loop may be comprised of sections of 26-, 24-, 22-, and 19-gauge conductors. The smaller gauges are usually employed close to the exchanges where the cables are larger and frequently in underground ducts. In the case of underground construction, the smaller gauge cables are an important factor in the conservation of duct space. On very long lines full loading is used to control transmission.

In urban areas, the larger cables are

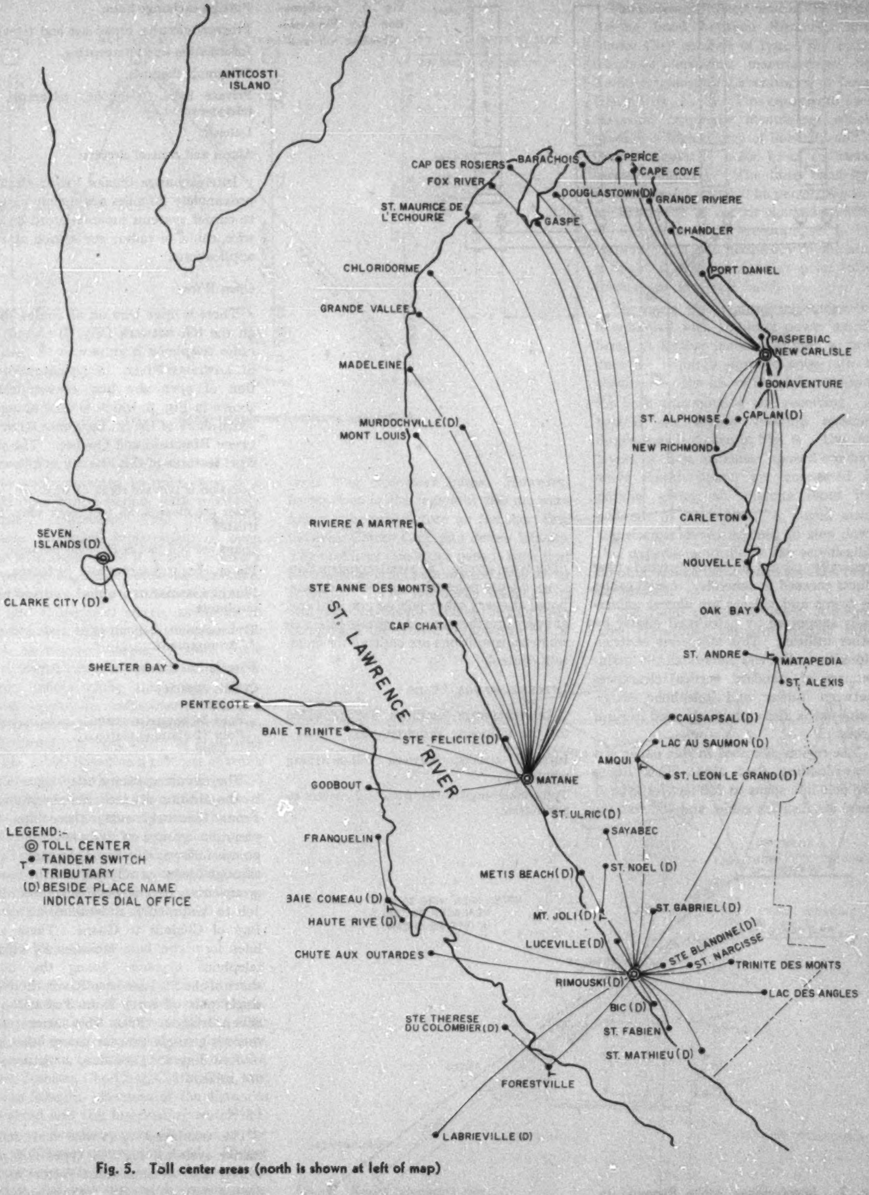

ANTICOSTI ISLAND
CAP DES ROSIERS
BARACHOIS
PERCE
CAPE COVE
FOX RIVER
DOUGLASTOWN(D)
GRANDE RIVIERE
ST. MAURICE DE L'ECHOURIE
GASPE
CHANDLER
CHLORIDORME
PORT DANIEL
GRANDE VALLEE
PASPEBIAC
NEW CARLISLE
MADELEINE
BONAVENTURE
ST. ALPHONSE
CAPLAN (D)
MURDOCHVILLE (D)
MONT LOUIS
NEW RICHMOND
SEVEN ISLANDS (D)
CARLETON
RIVIERE A MARTRE
NOUVELLE
CLARKE CITY (D)
STE ANNE DES MONTS
OAK BAY
CAP CHAT
ST. LAWRENCE RIVER
ST. ANDRE
MATAPEDIA
SHELTER BAY
ST. ALEXIS
PENTECOTE
CAUSAPSAL (D)
BAIE TRINITE
STE FELICITE (D)
LAC AU SAUMON (D)
AMQUI
ST. LEON LE GRAND (D)
GODBOUT
MATANE
ST. ULRIC (D)
SAYABEC
LEGEND:-
TOLL CENTER
TANDEM SWITCH
TRIBUTARY
(D) BESIDE PLACE NAME INDICATES DIAL OFFICE
FRANQUELIN
METIS BEACH (D)
ST. NOEL (D)
BAIE COMEAU (D)
MT. JOLI (D)
ST. GABRIEL (D)
HAUTE RIVE (D)
STE BLANDINE (D)
LUCEVILLE (D)
ST. NARCISSE
CHUTE AUX OUTARDES
TRINITE DES MONTS
RIMOUSKI (D)
LAC DES ANGLES
BIC (D)
STE THERESE DU COLOMBIER (D)
ST. FABIEN
ST. MATHIEU (D)
FORESTVILLE
LABRIEVILLE (D)

Fig. 5. Toll center areas (north is shown at left of map)

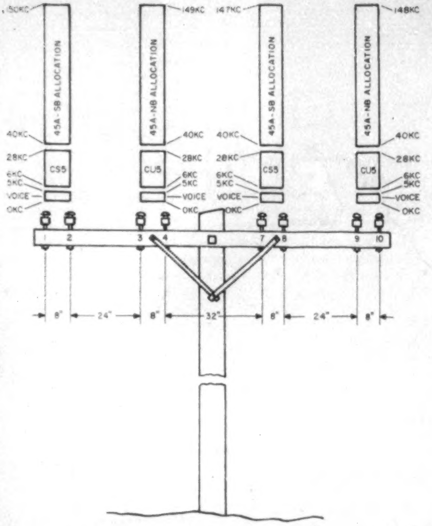


Fig. 6. Configuration of Rimouski-Quebec toll lead

generally placed in underground fiber ducts encased in concrete. Aerial cables in towns and villages are almost universally supported by poles used jointly by other utilities. This minimizes obstructions to traffic and facilitates the maintenance of standard vertical clearances between power and telephone wires. Joint use is also extensively used in rural areas.

The region discussed in this paper is a heavy-loading sleet area. This limits the pole-line spans to 250 feet for type *B* rural distribution cable, and 350 feet for open-wire (grade *135* steel) construction. On joint-use poles, the open wire is transposed at every other pole on nonjoint use, at every fourth span. At certain intervals some transpositions are omitted for cross-talk control.

Interexchange Lines

Interexchange facilities are provided for the following requirements:

Intertoll trunks, between toll-switching points.

Toll-connecting trunks, from toll centers to tributaries.

Foreign-exchange lines.

Program circuits, broadcast and television.

Information and intercepting.

Telegraph channels.

Private lines, telephone, telegraph, and teletypewriter.

Defence.

Alarm and control circuits.

Interexchange trunks longer than approximately 20 miles are mainly assigned to carrier systems super-imposed on open wire, cable, or radio; see section on radio applications.

Open Wire

There is open wire on all routes shown on the toll network (Fig. 2) except that radio telephone is employed to span the St. Lawrence River. A typical configuration of open-wire line construction is shown in Fig. 6, which is used along the south shore of the St. Lawrence River between Rimouski and Quebec. The principal features of this line are as follows:

Location is private right of way.

Poles are class *4*, 25 feet, jack pine, fully treated.

Spans are 165 feet or 32 poles per mile.

Pin spacing is 8 inches and 24 inches.

Pins and insulators are steel pins and pyrex insulators.

Transposition is point-type and mid-span *J5* arrangement.

Wire is 128-mil hard-drawn copper.

Circuit capacity	
Physical	4
Four 3-channel systems	12
Four 12-channel systems	48
Total	64

The circuit capacity may be doubled by the addition of a secoud 8-pin crossarm. From Rimouski east, there are two phantom groups of copper wire. One group follows the St. Lawrence River through Matane to Gaspé, and the second group cuts across the peninsula from Mont Joli to Matapedia, and thence along the Bay of Chaleur to Gaspé. These wire lines form the base facilities for carrier telephone systems. Along the north shore of the St. Lawrence River, there is a single pair of wires from Tadoussac to Seven Islands. From this latter point, there is a single wire for morse telegraph to Red Bay in Labrador, a distance of 641 miles.

Cable

The combined open-wire and cable-carrier systems, such as types *ON* and *45BN*, are an important factor in reducing the cost of cable for interexchange trunks. These systems can be run in

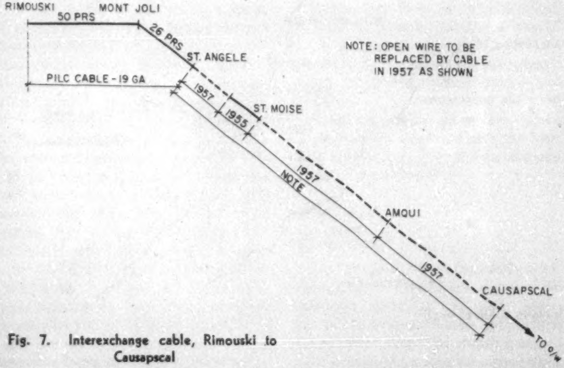
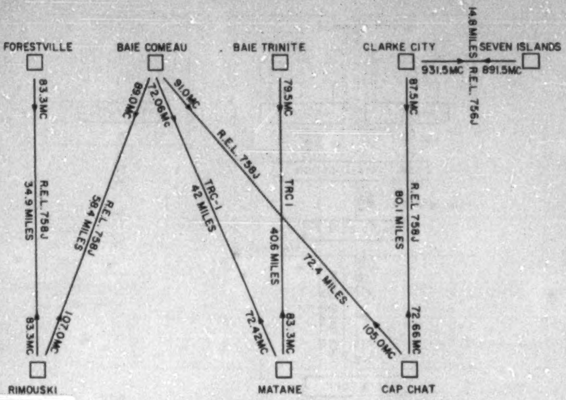


Fig. 7. Interexchange cable, Rimouski to Causapscal



FORESTVILLE
BAIE COMEAU
BAIE TRINITE
CLARKE CITY
SEVEN ISLANDS
14.8 MILES
R.E.L. 756J
931.5MC
891.5MC
83.3MC
R.E.L. 758J
34.9 MILES
83.3MC
89.0MC
72.06MC
91.0MC
R.E.L. 758J
58.4 MILES
107.0MC
TRC-1
42 MILES
72.42MC
R.E.L. 758J
72.4 MILES
105.0MC
79.5MC
TRC 1
40.6 MILES
83.3MC
87.5MC
R.E.L. 758J
80.1 MILES
72.66MC
RIMOUSKI
MATANE
CAP CHAT

Fig. 8. Radio-telephone frequency assignments

cable and connected to open wire at a cable open-wire junction, with relatively simple equipment. This flexibility greatly facilitates replacement of open wire with cable since it is not necessary to replace an entire route at one time. This feature is being used between Rimouski and Causapscal where, as shown in Fig. 7, open wire is being replaced by cable in planned stages. Loaded cable pairs in this cable are used for trunks between offices along the route. Combined open-wire cable-carrier systems will be employed for through-circuit requirements. A cost study of open wire versus cable, based on a forecast of trunk requirements during the period from 1955 to 1960, showed that the capital cost of the cable arrangement would be substantially less than an equivalent open-wire plan. Moreover, the ultimate capacity of the cable was greater, and cable has definite advantages in this heavy-loading sleet area.

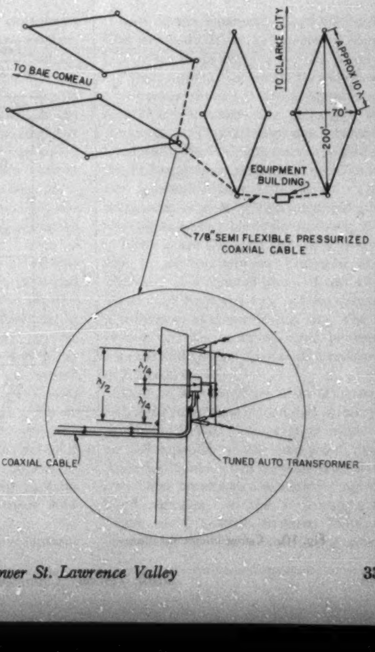
Radio Telephone

The radio telephone systems shown in Fig. 2 are used to provide communications across the St. Lawrence River. The principal system is comprised of four links operating in tandem between Rimouski, Baie Comeau, Cap Chat, Clarke City, and Seven Islands. Because of the distances involved and the intervening water, frequency allocations were obtained in the 72- to 108-mc (megacycle) band. Higher frequencies which are in more general use today for point-to-point communications were considered unsatisfactory for these long overwater paths. However, frequencies in the region of 900 mc were found to be satisfactory for the short link between Clarke City and Seven Islands. The profile of this latter path is such that interference is not experienced from water reflections. The frequency assignments and path lengths for each of the four links comprising the system is shown in Fig. 8.

The radio equipment used in the 72- to 108-mc band between Rimouski and Clarke City consists of type *758J* radio telephone terminals manufactured by Radio Engineering Laboratories of Long Island City, N. Y. This equipment uses Serrasoid frequency modulators which produce a high degree of linearity and a peak deviation of ±500 kc of the radio carrier frequency. The base band extends from 2 kc to 300 kc permitting the application of 72 carrier channels spaced at 4-kc intervals. Pre-emphasis is incorporated in the frequency modulator to yield a uniform signal-to-noise ratio throughout the base band.

The basic *758J* transmitter equipment is equipped with 100-watt power amplifiers. A further increase in gain to produce the required signal-to-noise ratio is obtained by the use of rhombic antennas for both transmitting and receiving. A typical site layout showing antenna dimensions is shown in Fig. 9. The two layers of these antennas, spaced one-half wave length apart, are connected in parallel giving an antenna input impedance of 300 ohms. A tuned auto-transformer device is used to step down this impedance to match the semiflexible 50-ohm coaxial cable leading to the radio equipment.

The radio system between Rimouski and Forestville employs similar radio



TO BAIE COMEAU
TO CLARKE CITY
APPROX 10λ
70'
200'
EQUIPMENT BUILDING
7/8" SEMI FLEXIBLE PRESSURIZED COAXIAL CABLE
λ/4
λ/2
λ/4
COAXIAL CABLE
TUNED AUTO TRANSFORMER

Fig. 9. Radio telephone site layout at Cap Chat

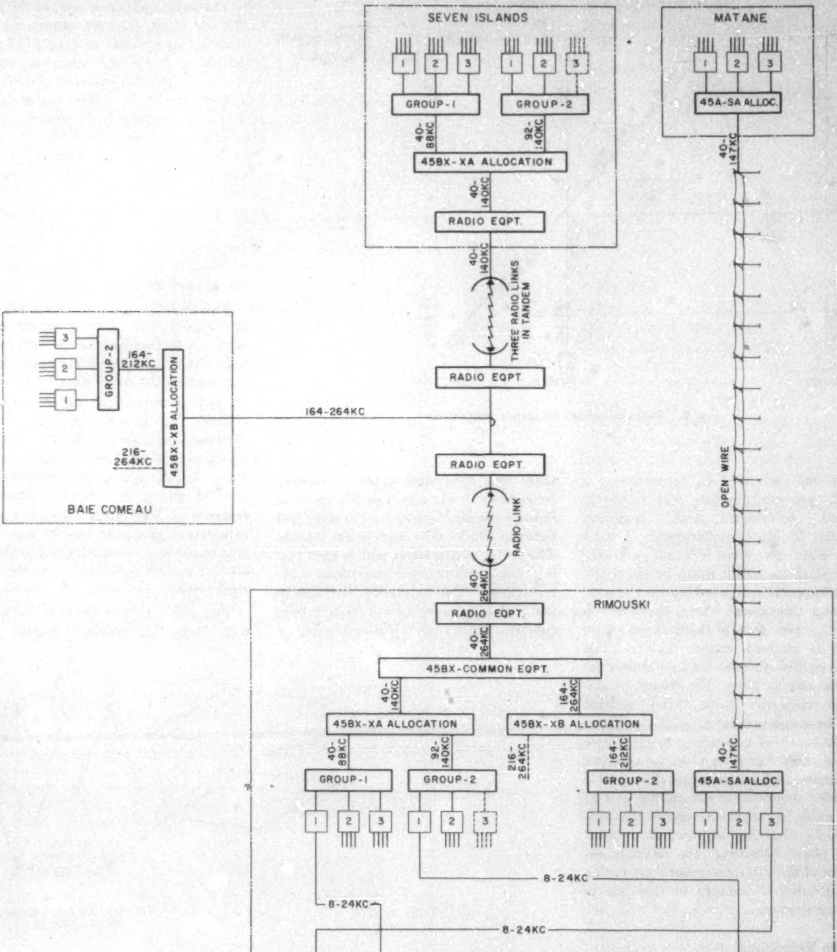

SEVEN ISLANDS
1
2
3
1
2
3
GROUP-1
GROUP-2
40-88KC
92-140KC
45BX-XA ALLOCATION
40-140KC
RADIO EQPT.
40-140KC
THREE RADIO LINKS IN TANDEM
RADIO EQPT.
MATANE
1
2
3
45A-SA ALLOC.
40-147KC
OPEN WIRE
BAIE COMEAU
3
2
1
GROUP-2
164-212KC
45BX-XB ALLOCATION
216-264KC
164-264KC
RADIO EQPT.
RADIO LINK
40-264KC
RADIO EQPT.
RIMOUSKI
40-264KC
45BX-COMMON EQPT.
40-140KC
164-264KC
45BX-XA ALLOCATION
45BX-XB ALLOCATION
40-88KC
92-140KC
216-264KC
164-212KC
40-147KC
GROUP-1
GROUP-2
GROUP-2
45A-SA ALLOC.
1
2
3
8-24KC
8-24KC
8-24KC
45A-SA ALLOC.
45A-NB ALLOC.
45A-SB ALLOC.
45A-NA ALLOC.
40-147KC
40-149KC
40-150KC
40-148KC
TO QUEBEC SEE FIG. 6

Fig. 10. Carrier interconnections

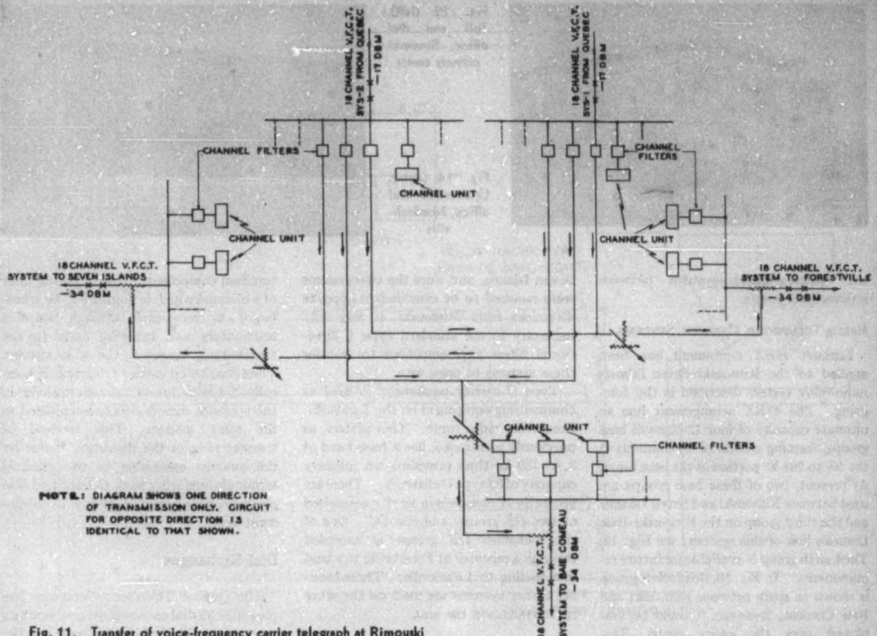

18 CHANNEL V.F.C.T. SYS-2 FROM QUEBEC
−17 DBM
18 CHANNEL V.F.C.T. SYS-1 FROM QUEBEC
−17 DBM
CHANNEL FILTERS
CHANNEL FILTERS
CHANNEL UNIT
CHANNEL UNIT
CHANNEL UNIT
18 CHANNEL V.F.C.T. SYSTEM TO SEVEN ISLANDS
−34 DBM
18 CHANNEL V.F.C.T. SYSTEM TO FORESTVILLE
−34 DBM
CHANNEL UNIT
CHANNEL FILTERS
NOTE: DIAGRAM SHOWS ONE DIRECTION OF TRANSMISSION ONLY. CIRCUIT FOR OPPOSITE DIRECTION IS IDENTICAL TO THAT SHOWN.
18 CHANNEL V.F.C.T. SYSTEM TO BAIE COMEAU
−34 DBM

Fig. 11. Transfer of voice-frequency carrier telegraph at Rimouski

equipment except it has a narrower base band, −2 kc to −160 kc. However, the path length is 34.9 miles shorter, and power amplifiers for the transmitters are not required. Lower-gain corner-reflector antennas were found to be adequate for this link. The other systems in the region have a capacity of only three voice channels and are of less interest here.

The radio equipment is operated by a 115-volt supply of alternating current. Commercial power is available at all sites. However, to maintain continuity of service, each station is equipped with a stand-by engine generator set. This unit is automatically started and takes over the load in case of failure of the regular power supply.

Carrier Telephone Equipment

Open-Wire Carrier Telephone Systems

The application of carrier to the main toll route between Rimouski and Quebec is illustrated in Fig. 6. Each pair of wires carries a 3-channel type *C* system and a 12-channel type *45A* system. These systems are equipped with automatic regulators to compensate for changes in line attenuation caused by temperature, wet and dry weather, etc. The 12-channel systems also have intermediate repeaters or amplifiers spaced at 60-mile intervals. Part of these systems is extended beyond Rimouski on land lines and radio telephone. The method of transferring groups of carrier frequencies from wire lines to radio is shown in Fig. 10.

Types *C* and *O* carrier equipment are employed on the toll lead between Rimouski and New Carlisle. Two *C* systems and an *OB* group branch off this route at Matapedia to Campbellton and at other points in the Province of New Brunswick; see Fig. 2. On this route *O*-type repeaters are used at Amqui, Matapedia and Carleton. An *OD* group is dropped off at Amqui to provide Rimouski-Amqui trunks. It is planned to convert the *O* equipment to *ON*-type cable version between Rimouski and Causapscal in accordance with the cable construction program shown in Fig. 7. Three-channel carrier telephone systems are used on the route between New Carlisle and Gaspé.

Types *C*, *O*, and *45A* carrier equipment is used on the Rimouski-Gaspé toll lead. The Rimouski–Mont Joli section is common to the route described in the preceding paragraph. This created a complication because from a crosstalk standpoint type *O* is not compatible with type *45* equipment. This difficulty was overcome temporarily by leasing frequency space for the *45A* equipment on another pole line in the Rimouski–Mont Joli section. The permanent solution will be the transfer of the *O* systems to cable as shown in Fig. 7.

There is one *C* system and one *45A* system in service between Rimouski and Matane. A 4-channel group of this *45A* system is transferred to a Quebec system at Rimouski as shown in Fig. 10. The *C* system on this route operates between Matane and Quebec City with a repeater at Rimouski.

East of Matane there is an *OB* group operating to an unattended dial exchange at Murdochville. This system employs an *OB* repeater at the mid-point, Ste. Anne des Monts. Dialing is accomplished over these trunks by the built in out-of-band signaling of the *O* equipment. There are *C* systems between Matane, Ste. Anne des Monts, and Gaspé along

Fig. 12 (left). Toll and dial office, Rimouski primary center

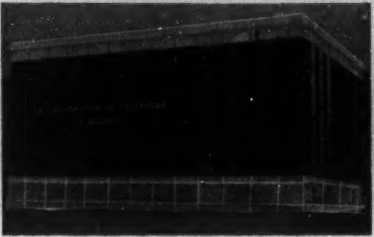


Fig. 14 (right). Unattended dial office, Murdochville

with single-channel systems between intermediate points.

Radio-Telephone Carrier Systems

Lenkurt *45BX* equipment has been applied to the Rimouski–Seven Islands radio-relay system described in the foregoing. The *45BX* arrangement has an ultimate capacity of four 12-channel base groups, forming a total of 48 channels in the 40- to 264-kc portion of the base band. At present, two of these base groups are used between Rimouski and Seven Islands and the third group on the Rimouski-Baie Comeau link of this system; see Fig. 10. The fourth group is available for future requirements. In Fig. 10, this fourth group is shown as spare between Rimouski and Baie Comeau, however, it could be employed over the entire route. This flexibility is obtained with the use of carrier hybrids for dropping off groups of frequencies at intermediate points. These hybrids pass all frequencies in the base band.

There are also two *C* systems in operation on the Rimouski–Seven Islands radio system. East and west terminals are used at the Seven Islands terminal to provide a suitable frequency arrangement for the operation of six channels under the type *45BX* equipment. At the Rimouski radio terminal these two 3-channel systems are transferred to wire lines. Because east and west terminals are used at Seven Islands, and since the two systems were required to be extended in opposite directions from Rimouski, it was only necessary to use standard type *C* directional filters and amplifiers to transfer these systems to open wire.

Type *O* carrier equipment is used as channelizing equipment on the Rimouski-Forestville radio route. This system, as previously mentioned, has a base band of 2 to 160 kc thus providing an ultimate capacity of 32 type *O* channels. There are presently 12 channels in service comprised of two *OB* groups and one *OC*. One of the 4-channel *OB* groups is extended through a repeater at Forestville to a land line leading to Labrieville. Three-channel carrier systems are used on the other radio systems in the area.

Voice-Frequency Carrier Telegraph

Voice-frequency carrier telegraph systems are operated over channels of carrier telephone equipment, both on wire lines and radio telephone. There are five 18-channel systems operating out of Rimouski, two to Quebec City and single systems to Forestville, Baie Comeau, and Seven Islands. These amplitude-modulated carrier telegraph systems use on-off keyed voice-frequency tones for transmission of teletypewriter service. Fig. 11 shows an arrangement used at Rimouski for transferring telegraph channels from one system to another without the use of terminal equipment. The receiving filter of a channel which is required to be transferred is connected through suitable attenuators and bridging pads to the transmitting branch of the other system. Voice-frequency carrier telegraph is basically a 4-wire circuit, and the transfer in the opposite direction is accomplished in the same manner. This method of transfer reduces the distortion factor by the amount assessable to two channel terminals operating back to back and also produces a substantial saving in equipment costs.

Dial Exchanges

The Québec Téléphone Company has presently 39 dial exchanges throughout its service area, 23 of which are located in the area discussed in this paper. Thus it is desirable to standardize, in as far as possible, types of equipment, layouts, and buildings. Uniformity conserves engineering, and installation time also simplifies maintenance and future additions.

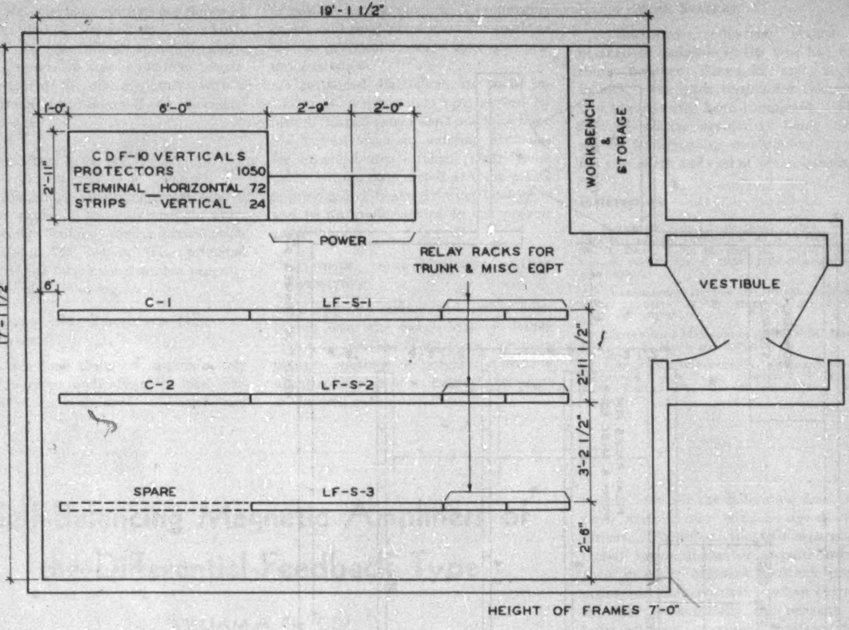
Equipment

The following is an outline of the dial equipment used in the area:

General.........	step-by-step; terminal-per-station.
Line finders.....	200 point.
Selectors........	digit-absorbing.

Fig. 13 (left). Toll and dial office, Seven Islands toll center

Fig. 15 (right). Typical small community dial exchange building


19'-1 1/2"
17'-1 1/2"
1'-0"
6'-0"
2'-9"
2'-0"
2'-11"
CDF-10 VERTICALS
PROTECTORS 1050
TERMINAL STRIPS HORIZONTAL 72
VERTICAL 24
POWER
WORKBENCH & STORAGE
VESTIBULE
RELAY RACKS FOR TRUNK & MISC EQPT
6"
C-1
LF-S-1
C-2
LF-S-2
SPARE
LF-S-3
2'-11 1/2"
3'-2 1/2"
2'-6"
HEIGHT OF FRAMES 7'-0"

SWITCH FRAME ARRANGEMENTS

FRAME	CAPACITY		
	LINE FINDER UNITS (200 LINES, 16 SWITCHES)	SEL SHELVES 10 SWITCHES EACH	CONN GROUPS 100 TERMINALS 11 SW. EACH
L.F. FR NO 1	1	2	
L.F. FR NO 2	1	2	
L.F. FR NO 3	1	2	
CONN FR NO 1			4
CONN FR NO 2			4
TOTAL	3 (600 LINES)	6 (60 SW.)	8 (800 TER.)

Fig. 16. Typical dial office floor plan, 800 terminals

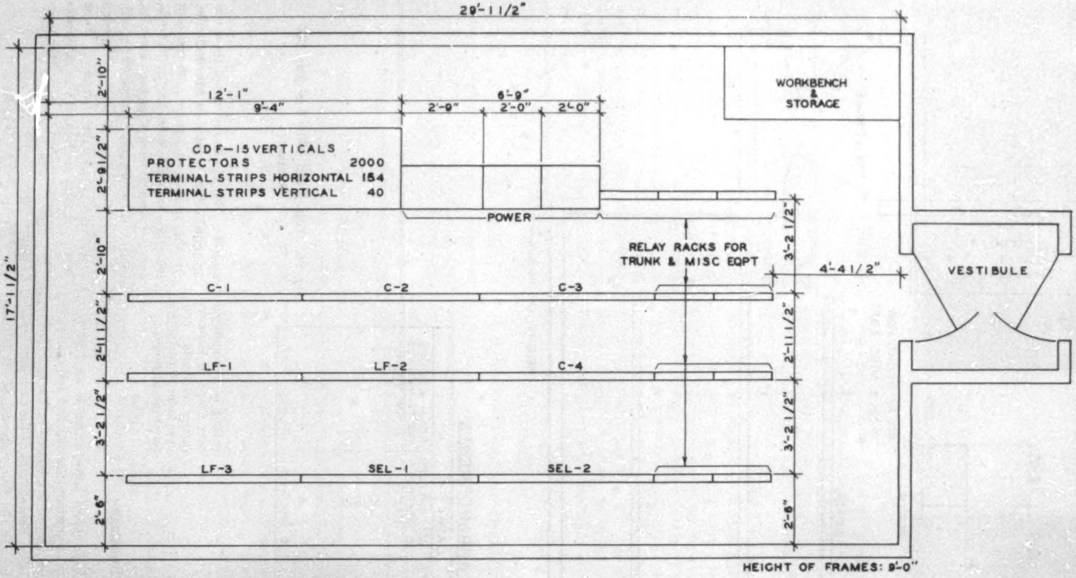
Connectors	1- and 2-party; rotary hunting for PBX trunks and other consecutive number stations; 10-party for rural service.
Reverting calls	2-party lines, 1191; 10-party, directory listing.
Intercepting	by operator at toll center.
Signaling	busy flashes and tones.
Pay stations	prepayment with exchanges close to toll centers; semi prepayment at other points.
Trunks-to-toll Centers	toll recording and toll switching where dial exchange in same building as toll board. Two-way master office trunks at other points. Signalling is accomplished over *E* and *M* leads on carrier channels, and over loops or composite on physical circuits.

Buildings

Fireproof construction is used at toll centers where the dial and toll-switching equipment are housed in the same building. Photographs of the exchange buildings at the Rimouski and Seven Islands toll centers are shown in Figs. 12 and 13. At unattended exchanges, fireproof buildings are also specified where the ultimate capacity is forecast to be more than 800 stations. Fig. 14 shows one of these unattended offices located at Murdochville. At the smaller exchanges, fire-resistant materials are used to reduce fire hazards, e.g., outside sheathing of asbestos shingles, interior walls and ceiling of plaster board. Fig. 15 shows a typical small unattended dial office. As a further precaution, these small buildings are located at a minimum distance of 40 feet from any other nonfireproof structure.

FLOOR PLAN LAYOUT

Standard floor plan layouts are used at all unattended dial exchanges. Layouts for offices with ultimate capacities for 800


29'-11 1/2"
17'-11/2"
2'-10"
12'-1"
9'-4"
6'-9"
2'-9"
2'-0"
2'-0"
2'-9 1/2"
CDF-15 VERTICALS
PROTECTORS 2000
TERMINAL STRIPS HORIZONTAL 154
TERMINAL STRIPS VERTICAL 40
POWER
WORKBENCH & STORAGE
RELAY RACKS FOR TRUNK & MISC EQPT
2'-10"
C-1
C-2
C-3
2'-11 1/2"
LF-1
LF-2
C-4
3'-2 1/2"
LF-3
SEL-1
SEL-2
2'-6"
3'-2 1/2"
4'-4 1/2"
2'-11 1/2"
3'-2 1/2"
2'-6"
VESTIBULE
HEIGHT OF FRAMES: 9'-0"

SWITCH FRAME ARRANGEMENTS

FRAME	CAPACITY		
	LINE FINDER UNITS (200 LINES, 16 SWITCHES)	SEL SHELVES 10 SWITCHES EACH	CONN GROUPS 100 TERMINALS 11 SW. EACH
L.F. FR NO 1	2	1	
L.F. FR NO 2	1	1	
L.F. FR NO 3	1	3	
SEL FR NO 1		6	
SEL FR NO 2		6	
CONN FR NO 1			5
CONN FR NO 2			5
CONN FR NO 3			5
CONN FR NO 4			2
TOTAL	5 (1000 LINES)	17 (170 SW)	17 (1700 TER)

Fig. 17. Typical dial office floor plan, 1,700 terminals

and 1,700 telephone stations are shown in Figs. 16 and 17 respectively. Standardization is a more difficult accomplishment at toll centers because it involves the coordination of the dial equipment with a wide range of terminal toll and associated equipment.

Future Plans

In concluding, it is appropriate to give a brief outline of plans to meet the ever-increasing demand for communication services in the region. The principal items of the future construction program include the following.

DEFERRED APPLICATIONS FOR LOCAL SERVICE

At this time there are approximately 3,200 deferred applications for local telephone service in the area. Additions and extensions for the purpose of exchanging cables and rural plant are to be made as soon as practical to serve these prospective customers.

Unattended dial offices are to be installed in communities now served by heavily loaded group lines and which are too distant from an existing exchange for extended area service. Distribution cables will be constructed at these points to provide facilities to serve new customers and to up grade service to our present subscribers.

CONVERSION FROM MANUAL TO DIAL OPERATION

To improve service and derive fuller benefit from the direct-distance dialing plan, the Québec Téléphone Company plans to continue its program of mechanizing a number of its manual exchanges in the area each year.

RADIO-RELAY SYSTEMS

A microwave radio-relay system is planned to supplement the wire-line toll route between Rimouski and Seven Islands. The path testing for this system has recently been completed. The forward scatter system is being considered for extending communications to the area north and east of Seven Islands.

References

1. RAILROAD COMMUNICATIONS USING HIGH-POWER CARRIER TECHNIQUE, A. K. Hansen, W. W. H. Dean, F. G. Mullins, Jr. *AIEE Transactions*, vol. 72, pt. I, 1953 (Jan. 1954 section), pp. 785–91.

2. TRANSMISSION DESIGN OF INTERTOLL TELEPHONE TRUNKS, H. R. Huntley. *Ibid.*, Nov. 1953, pp. 670–76.

3. NATION-WIDE NUMBERING PLAN, W. H. Nunn. *Ibid.*, vol. 71, pt. I, pp. 257–60.

4. SIMPLIFIED TRANSMISSION ENGINEERING IN EXCHANGE CABLE PLANT DESIGN, L. B. Bogan, K. D. Young. *Ibid.*, vol. 73, pt. I, Nov. 1954, pp. 498–502.

Self-Balancing Magnetic Amplifiers of the Differential-Feedback Type

WILLIAM A. GEYGER
MEMBER AIEE

Synopsis: This paper is concerned with the application of differential-feedback circuitry to push-pull self-balancing magnetic amplifiers in order to obtain extreme stability and freedom from zero drift with a minimum of core matching and component selection. It describes a 60-cycle magnetic instrument amplifier particularly well suited for extending the range of application of a standard-type moving-coil ink recorder and gives the results of various measurements and actual performance characteristics. This amplifier acts as an operational amplifier and is suitable for various applications in instrumentation, telemetering, automatic control, and analog-computer technique.

IN push-pull-type magnetic amplifiers with two single-ended external-feedback or self-saturating circuits in back-to-back connection, each circuit has its own positive feedback component which reduces or nearly nullifies the intrinsic negative feedback of the ordinary saturable reactor (transductor). As a result, the individual quiescent currents of these circuits vary considerably with large changes in supply voltage, frequency, and ambient temperature. Therefore, the characteristics of the magnetic core and rectifier elements must be carefully matched over a large part of the working range to avoid excessive asymmetry zero-drift errors. This necessity becomes particularly important in low-level d-c signal applications in the realm of instrumentation, control engineering, and analog-computer technique.

With regard to this situation, the author[1,2] was forced, early in 1939, to abandon the afore-mentioned push-pull arrangement. In order to reduce the severe requirements on matching of the magnetic cores and rectifiers, an entirely different form of push-pull arrangement with two separately biased ordinary saturable-reactor circuits and differential feedback was introduced at that time.[3,4] In this case, the resultant positive feedback component of the magnetic amplifier is proportional to the actual difference between the individual load currents and is zero with no-signal conditions. Therefore, the two separately biased saturable-reactor circuits retain their inherent negative feedback properties, and their resultant output current is then substantially independent of large changes in supply voltage, frequency, and ambient temperature.

The great importance of this basic concept was stressed in 1948 by C. S. Hudson[5] who pointed out that in "the self-excited type of transductor [with two single-ended circuits each having its own positive feedback component] the cross-over point on the axis wanders, and in using two back-to-back it is necessary to match their characteristics over a large part of the range. whereas, if the transductor is biased to the working point and amplification attained by true positive feedback, better stability with a smaller zero drift is obtained." In England, push-pull-type magnetic amplifiers utilizing bias excitation[6] with differential feedback have been modified and applied in various forms.[7,8]

Another approach to the asymmetry zero-drift problem consists in applying magnetic amplifiers of the self-balancing potentiometer type[9,10] which use a special combination of external or internal magnetic positive feedback and electric (galvanic) negative feedback. In accordance with the basic principle of these self-balancing magnetic amplifiers, positive external feedback or self-saturation produces an effectively infinite gain that is highly degenerated by 100% (per

---

Paper 57-92, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted September 17, 1956; made available for printing December 5, 1956.

WILLIAM A. GEYGER is with the U. S. Naval Ordnance Laboratory, Silver Spring, Md.

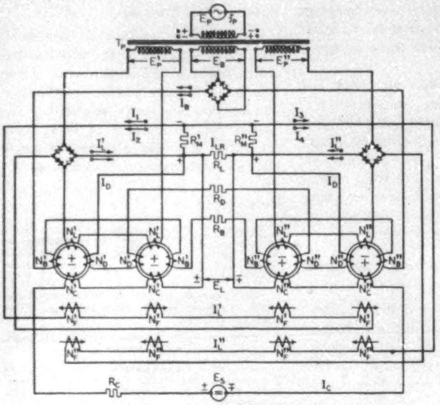


Fig. 1. Circuit diagram of basic voltage-controlled self-balancing magnetic amplifier of differential-feedback type

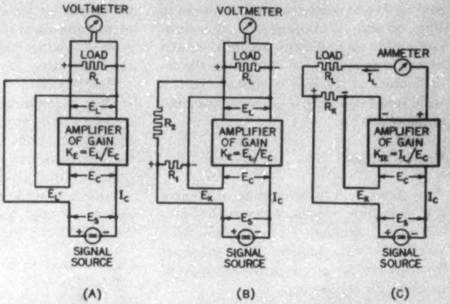


Fig. 2. Block diagrams of various forms of voltage-controlled self-balancing magnetic amplifiers of differential-feedback type

A—$E_S = E_L$
B—$E_S = E_L\beta = E_L R_1/(R_1+R_2)$
C—$E_S = I_L R_K$

cent) negative feedback resistance-coupled into the control-circuit loop. The drastic reduction of zero drift in self-balancing magnetic amplifiers results from the combination of the dynamic gain-increasing effect of magnetic positive feedback and the drift-reducing effect of electric negative feedback.[11,12]

Technical details and successful application of push-pull self-balancing magnetic amplifiers for air-borne servo control systems and aircraft instrumentation were described recently by R. M. Hubbard[13] and R. Cockrell. In these arrangements, two single-ended self-saturating circuits in back-to-back connection are used in combination with a center-tapped power-supply transformer and a mixing-resistor network. The report[13] makes it evident that these amplifiers meet the stringent environmental requirements of aircraft operation, and lend themselves readily to quantity-production techniques with a minimum of core matching and component selection.

A novel technique for obtaining extreme stability and freedom from zero drift consists in applying differential-feedback circuitry to push-pull self-balancing magnetic amplifiers. In this way, it is possible to combine the specific advantages of both artifices and to obtain optimum performance. This is, of course, particularly significant for low-level d-c signal applications and for solutions of those special problems where extremely small values of linearity error and zero drift represent prime requirements.

Voltage-Controlled Self-Balancing Magnetic Amplifiers of the Differential-Feedback Type

Operating Principle of the Basic Circuit

In the first example of a voltage-balance circuit, as shown in Fig. 1, the polarity-reversible signal voltage $\pm E_S$ is balanced by means of several voltage components which are supplied in time sequence by the polarity-reversible load voltage $E_L = I_{LR}R_L$, and by additional voltages induced in the control-winding elements from nonfired cores during their gating half-cycles. As a result the balanced control circuit demands practically no energy from the d-c signal source, and the effective input resistance is about 100 to 1,000 times higher than the actual input-circuit resistance.

The self-balancing circuit, Fig. 1, was derived from the original system (see Figs. 28–31, reference 4) by making the following changes:

1. Two groups of additional d-c control windings N_F' and N_F'', acting as differential-feedback windings, are provided for producing an effectively infinite internal gain of the amplifier.
2. In order to produce electric (galvanic) negative feedback, the terminals of load resistor R_L are series-connected with the input circuit in such a way that the load voltage E_L opposes the signal voltage E_S.
3. Feedback windings N_D' and N_D'' are provided for producing a small amount of additional (positive or negative) magnetic feedback which is adjusted by means of the variable regeneration-control resistor R_D in such a way that the control-winding current I_C is substantially reduced to zero.

If I_C is very small in comparison with I_{LR}, the average value of the control-winding current will be

$$I_C = \frac{E_S - E_L}{R_C + R_L} = \frac{E_S}{R_C + R_L} - \frac{E_L}{R_C + R_L} = I_S - I_K \tag{1}$$

where E_S is the electromotive force of the signal source, and R_C is the total resistance of the control circuit between the terminals of load resistor R_L, including the sum of the resistances of the four control-winding units with series resistor, and the internal resistance of the signal source. The term $I_S = E_S/R_C + R_L$ is the signal-current component and $I_K = E_L/R_C + R_L$ is the compensating-current component.

If the regeneration-control circuit N_D'-$N_D''R_D$ is disconnected from the terminals of load resistor R_L, then the ampere-turns of the two saturable-reactor systems with differential feedback are

$$I_L'N_L' = I_BN_B' + I_CN_C' + I_L'N_F' - I_L''N_F'' \tag{2}$$

and

$$I_L''N_L'' = I_BN_B'' - I_CN_C'' - I_L'N_F' + I_L''N_F'' \tag{3}$$

With $N_L' = N_L'' = N_L$, $N_B' = N_B'' = N_B$, and $N_C' = N_C'' = N_C$

$$(I_L' - I_L'')N_L = 2I_CN_C + 2(I_L' - I_L'')N_F \tag{4}$$

Thus, the average value of control-winding current may be expressed by

$$I_C = I_{LR}\left(\frac{2R_M + R_L}{R_M}\right)\left(\frac{N_L - 2N_F}{2N_C}\right) \tag{5}$$

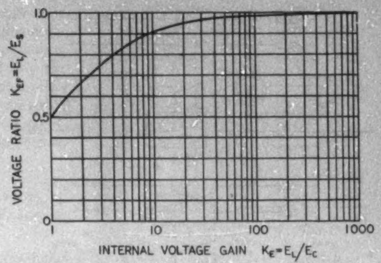


Fig. 3. External voltage gain $K_{EF}=E_L/E_S$ as a function of internal voltage gain $K_E=E_L/E_C$ of the amplifiers of Figs. 1 and 2(A)

Substituting from equation 1 into equation 5 gives

$$\frac{E_S-I_{LR}R_L}{R_C+R_L}=I_{LR}\left(\frac{2R_M+R_L}{R_M}\right)\left(\frac{N_L-2N_F}{2N_C}\right) \tag{6}$$

or

$$E_S=I_{LR}(R_C+R_L)\times \left[\frac{R_L}{R_C+R_L}+\left(\frac{2R_M+R_L}{R_M}\right)\left(\frac{N_L-2N_F}{2N_C}\right)\right] \tag{7}$$

Thus, with $N_F=N_L/2$

$$E_S=I_{LR}R_L=E_L \tag{8}$$

In this case, the control circuit is balanced, thus

$$I_C=I_S-I_K=0 \tag{9}$$

However, the condition $N_F=N_L/2$ which corresponds to critical regeneration (infinite internal gain of the magnetic amplifier[9,10]) involves the assumption of an ideal feedback-rectifier characteristic represented by rectifier-output currents I_L' and I_L'' having exactly the average values of the rectifier-input currents I_P', I_P''. In practice, however, departure of the feedback rectifiers from the ideal characteristic causes a certain reduction of current ratios I_L'/I_P' and I_L''/I_P'', so that actual differential-feedback ampere-turns $I_L'N_F'$ and $I_L''N_F''$ are reduced.

The additional feedback windings N_D' and N_D'' with variable series resistor R_D make it possible to compensate for imperfections of the saturable reactors and rectifiers by adjusting the additional positive magnetic feedback of these windings in such a way that the control-winding current $I_C=I_S-I_K$ is reduced to zero. Actually, the effective positive feedback ampere-turns will be equal to $(I_L'-I_L'')N_F+I_DN_D$. In the example of Fig. 1, $I_D=E_L/R_D$, where R_D is the total resistance of the regeneration-control circuit $N_D'N_D''R_D$.

Of course, it is also possible to make $N_F'=N_F''=N_F$ larger than $N_L/2$, and to use the windings N_D' and N_D'' for applying a proper amount of negative magnetic feedback to obtain the condition of infinite internal gain of the amplifier. In this case, the resultant positive feedback ampere-turns $(I_L'-I_L'')N_F-I_DN_D$ are adjusted so that $I_C=I_S-I_K$ is substantially reduced to zero.

Proper adjustment of regeneration-control resistor R_D may be achieved easily by means of a microammeter consisting of a lightspot galvanometer with low-resistance shunt. If R_D is adjusted under or over the point of critical regeneration while the input circuit is subjected to variations of signal voltage E_S, a control-winding current I_C will appear but will have a direction of current flow dependent upon the direction of actual misadjustment. With critical regeneration, I_C is in the order of $0.001\times I_{LR}$, and the signal voltage E_S has only a transient effect upon the magnetic amplifier.[9,10]

Properties and Modifications of the Basic Circuit

Fig. 2(A) is a block diagram of the basic circuit of Fig. 1, and Figs. 2(B) and 2(C) show important modifications of this circuit. In Fig. 1, and Fig. 2(A), the total amount of load voltage E_L is superposed upon the signal voltage E_S, and actual control voltage across the control windings N_C', N_C'' will be

$$E_C=E_S-E_L=I_CR_{CW} \tag{10}$$

where R_{CW} is the total resistance of the four series-connected control-winding units. The external voltage gain including the effects of the electric negative feedback is

$$K_{EF}=\frac{E_L}{E_S}=\frac{1}{1+\dfrac{1}{K_E}} \tag{11}$$

where $K_E=E_L/E_C$ is the internal voltage gain without negative feedback.

Fig. 3 which shows $K_{EF}=f(K_E)$ makes it evident that, when K_E is very much larger than unity, the voltage developed across the load-resistor R_L will be equal within very close limits to the signal voltage E_S. For example, if $K_E=1{,}000$, then load voltage E_L which may be indicated by a high-precision voltmeter will be less than the signal voltage E_S by only one part in 1,000, or less than 1/10 of 1.0%. Furthermore, because $K_E>>1$, it follows that actual voltage ratio E_L/E_S is substantially independent of changes in the amplifier itself, and that E_L will remain essentially equal to E_S over a wide range of variations in magnetic-core and feedback-rectifier characteristics, and large changes in supply voltage, frequency, and ambient temperature. Thus, if the internal voltage gain K_E should change, for example, from 1,000 to 500, the only effect would be to cause an error in the voltmeter reading of about 1/10 of 1.0%.

Evidently, in the basic circuit, Figs. 1 and 2(A), as a result of an extremely large amount of negative electric feedback, the external voltage ratio $K_{EF}=E_L/E_S$ has been reduced to 1.0. This ratio, however, will remain constant, if the internal voltage gain $K_E=E_L/E_C$ should change within wide limits. In order to increase K_{EF}, a fraction β of load voltage E_L may be superposed upon the signal voltage E_S; Fig. 2(B). Here, K_{EF} is equal within very close limits to the resistance ratio $(R_1+R_2)/R_1$ of the voltage divider across R_L, and remains constant if $K_E=E_L/E_C$ changes within wide limits.

Besides these facts, the circuits of Figs. 2(A) and 2(B) have advantageous properties since the effective internal resistance of the output circuit with negative feedback is $(1+K_E\beta)$ times smaller than its original value without feedback; and the effective input resistance with negative feedback is $(1+K_E\beta)$ times larger than its original value.

In the arrangement of Fig. 2(C), the negative-feedback voltage $E_K=I_LR_K$ is derived from a fixed resistor R_K carrying the load current I_L. In this case, the effective internal resistance of the output circuit with negative feedback is larger than its original value without feedback.

Current-Controlled Self-Balancing Magnetic Amplifiers of the Differential-Feedback Type

Fundamental Principle

Fig. 4 shows three different forms of negative-feedback networks for converting the basic voltage-balance system of Fig. 1 into a current-balance system

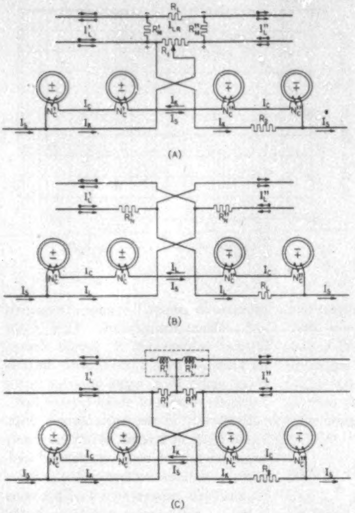


Fig. 4 (left). Negative feedback networks for converting the basic voltage-balance amplifier of Fig. 1 into a current-balance amplifier

A—Using series-mixing circuit
B—Using parallel-mixing circuit
C—Using twin-load circuit

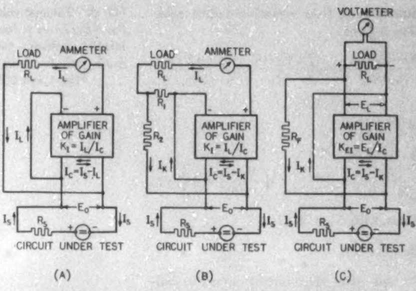


Fig. 5 (above). Block diagrams of various forms of current-controlled self-balancing magnetic amplifiers of differential-feedback type

A—$I_S = I_L$
B—$I_S = I_L\beta = I_L R_1/(R_1+R_2)$
C—$I_S = E_L/R_F$

by using a series-mixing circuit (A), a parallel-mixing circuit (B), or a twin-load circuit (C). In all cases, the signal current I_S of the circuit under test is balanced or "bucked out" by a compensating current I_K (A and C) or by the total load current I_L (B) so that actual voltage drop across the control windings N_C' and N_C'' is substantially reduced to zero.

PROPERTIES AND MODIFICATIONS OF CURRENT-CONTROLLED SELF-BALANCING MAGNETIC AMPLIFIERS

Fig. 5 shows block diagrams of current-controlled self-balancing magnetic amplifiers. In the circuit of Fig. 5(A), the total amount of load current I_L is superposed upon the signal current I_S, and the average value of I_L is equal within very close limits to I_S. In order to increase the current ratio I_L/I_S, a fraction β of load current I_L may be superposed upon the signal current I_S, as shown in Fig. 5(B). The feedback current $I_K = \beta I_L$ is derived from a current divider R_1R_2 which is series-connected with R_L. This is in contrast to the voltage-balance amplifier of Fig. 2(B), where a voltage divider is used to obtain the desired fraction β of load voltage E_L. In the arrangement of Fig. 5(C), the feedback current $I_K = E_L/R_F$ is provided by means of a fixed resistor R_F connected across the load-resistor terminals.

Technical Properties and Applications

The adaptability of differential-feedback type of self-balancing magnetic amplifiers to various problems in instrumentation, telemetering, automatic control, and analog-computer technique may be judged by considering their specific technical properties which are summarized in the following sections.

COMPARISON OF DIFFERENT FORMS OF NEGATIVE FEEDBACK NETWORKS

In all cases, the external gain of the system is substantially independent of large changes in the internal gain of the amplifier itself, and effective resistance values of both input and output circuit are influenced in a way which is advantageous with regard to practical applications. There are, however, basic differences between the specific properties of the various forms of these amplifiers, as illustrated in the block diagrams of Figs. 2 and 5. These differences result from the different forms of negative feedback networks and are summarized in the following statements:

1. In voltage-balance amplifiers, Figs. 1 and 2, the effective input resistance E_S/I_C is greatly increased, and no appreciable current I_C is drawn from the signal source. These amplifiers act as a voltmeter having a very high resistance.
2. In current-balance amplifiers, Figs. 4 and 5, the effective input resistance E_O/I_S is greatly reduced, and no appreciable voltage drop E_0 is introduced into the circuit under test. These amplifiers act as an ammeter having a very low resistance and a very small voltage drop across its terminals.
3. In voltage-output amplifiers, Figs. 1, 2(A), 2(B), and 5(C), which use voltage feedback, the effective internal resistance of the output circuit is greatly reduced. These amplifiers act as a low-impedance generator, and actual magnitude of output voltage E_L is substantially independent of large changes in load resistance R_L.
4. In current-output amplifiers, Figs. 2(C), 4, 5(A), and 5(B), which use current feedback, the effective internal resistance of the output circuit is greatly increased. These amplifiers act as a high-impedance generator, and actual magnitude of output current I_L is substantially independent of large changes in load resistance R_L.

TRANSIENT RESPONSE

The relationship between power gain, linearity error, and response time is such that if one property is accentuated, it is always at the expense of the other properties.[9] For example, although the external gain of a self-balancing magnetic amplifier is independent of the number of turns of the control windings N_C' and N_C'', it is generally desirable, in the interest of optimum linearity, to use as many turns as the required response time will permit. In this way the inevitable residue of compensating ampere-turns which must be supplied by the control circuit will require a minimum of current from that source, resulting in a minimum linearity error.[13] Furthermore, the use of the greatest possible number of turns of the control windings has the important added advantage of improving the stability of the amplifier by providing for a maximum of negative feedback ampere-turns.[13]

On the other hand, actual response time of self-balancing magnetic amplifiers may be reduced to its minimum value of one half-cycle of the power-supply frequency by introducing 100% negative feedback[14,15] without employing any rectifier components in the control circuit of the amplifier. Analysis of the voltage-balance form of this amplifier[16] reveals that half-cycle response is obtained if turns ratio $N_C/2N_L$ of external-feedback circuits (or the corresponding turns ratio of self-saturating circuits) is approximately equal to the resistance ratio R_K/R_O, where R_K is the actual resistance of the negative feedback resistor, and R_O is the total resistance of the output circuit including the load resistance, copper resistance of the output windings, and forward resistance of the rectifiers. Analysis of the current-balance form of the self-balancing magnetic amplifier reveals similar relationships.

In a previous paper,[9] it was stressed that the transient-response characteristics of self-balancing magnetic amplifiers may be altered materially by introducing differentiating or integrating circuit elements into the degeneration network of the amplifier itself. In this case, the amplifier does impose its special dynamic performance characteristics on the associated system with beneficial results.

Fig. 6(A), shows the possibility of reducing the response time of an indicating or recording moving-coil instrument in the output circuit of the amplifier by introducing a capacitor C_F with series resistor R_F into the negative feedback channel, to obtain a delayed negative feedback effect.[17] When the signal voltage E_S is suddenly applied to the input terminals of the amplifier, the instrument in its output circuit receives momentarily a very strong impulse because feedback voltage E_K and the fulfillment of voltage-balance condition are delayed as a result of the special transient performance of the feedback channel. In this way, with properly rated values of C_F and R_F, and with aperiodic damping conditions of the moving-coil system, its actual response time is considerably reduced. Similar effects are obtained with other forms of the amplifier, as shown in Figs. 2 and 5.

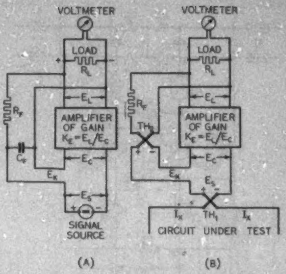


Fig. 6. Application of additional circuit elements in the negative-feedback channel of voltage-controlled self-balancing magnetic amplifiers of differential-feedback type

A—Introducing a capacitor C_F with series resistor R_F to obtain a delayed negative-feedback effect

B—Introducing a thermoconverter TH_2 to eliminate the effect of the nonlinear characteristic $E_S=f(I_X)$ of the thermoconverter TH_1 in the signal circuit

In accordance with common practice in the field of d-c amplifiers utilizing negative feedback, a capacitor or a mutual inductance may be introduced[9] into the degeneration network of the differential-feedback-type self-balancing magnetic amplifier to perform directly the integration or differentiation of a variable with respect to time.[1]

Application of Nonlinear Circuit Elements in the Negative-Feedback Channel

Although linearity of the output-input characteristic will be of prime importance in most applications, there are special cases where an exactly defined nonlinear relationship between signal voltage E_S or signal current I_S, and load voltage E_L or load current I_L may be desirable. For example, in Fig. 6(B), when measuring an alternating current I_X by means of a thermoconverter TH_1 having a square-law current-voltage characteristic

$$E_S=\text{constant}\times I_X^2 \tag{12}$$

it is advantageous to introduce a second thermoconverter TH_2 with a series resistor R_F into the negative-feedback channel.[17] In this case, the output-input characteristic of the amplifier will be altered in such a way that actual magnitude of load voltage E_L is proportional to the square root of the magnitude of signal voltage E_S

$$E_L=\text{constant}\times\sqrt{E_S} \tag{13}$$

Thus

$$E_L=\text{constant}\times I_X \tag{14}$$

Evidently, the basic principle of the arrangement of Fig. 6(B) consists in eliminating the effect of the nonlinear characteristic of thermoconverter TH_1 by introducing the equivalent nonlinear effect of thermoconverter TH_2 into the negative feedback channel of the amplifier to obtain a linear relationship between I_X and E_L. Of course, other types of nonlinear circuit elements may be used in such systems to modify the form of the output-input characteristic.

Special Applications

Differential-feedback-type self-balancing magnetic amplifiers are particularly suitable for extending the sensitivity of moving-coil ink recorders. It is possible to connect the input terminals of a current-balance amplifier, acting as a milliammeter without voltage drop, with the moving-coil circuit of a multirange instrument having numerous direct- and alternating-voltage and -current ranges without appreciably affecting the calibration of the instrument. In this case, the moving-coil ink recorder in the output circuit of the amplifier covers all the direct- and alternating-voltage and -current ranges of the multirange instrument.

These magnetic instrument amplifiers may also be used in combination with phase-sensitive rectifiers for the measurement of various a-c quantities, e.g., in a-c bridge and potentiometer networks. Voltage dividers and shunts may be provided to extend the alternating-voltage and -current ranges.

Nonelectric quantities are measured by means of conventional types of transducers representing special forms of variable resistors, inductances, mutual inductances, or capacitors, and providing the desired transition into an electrical value. When measuring and recording the photoelectric current produced by a photoelectric cell of the selenium barrier-layer type by means of a current-balance arrangement, shown in Figs. 4 and 5, the voltage across the terminals of the photoelectric cell is reduced to zero. As a result, the variation of the photoelectric current with respect to the illumination is practically linear.

In the fields of telemetering and automatic control, these magnetic amplifiers will undoubtedly find numerous interesting applications because of their distinct advantages over other d-c amplifier systems based upon the use of high-sensitivity galvanometers or vacuum tubes. With regard to the case of ap-

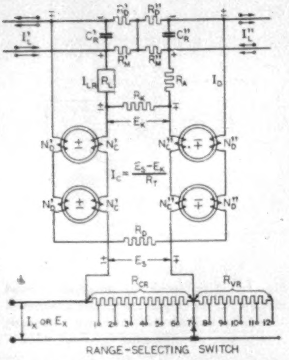


Fig. 7. Auxiliary resistance network with range-selecting switch for extending the range of application of a standard-type moving-coil ink recorder by means of a voltage-controlled self-balancing magnetic amplifier of differential-feedback type

plications in analog-computer technique, full advantage is taken of the specific properties, i.e., extreme stability, linearity error is about 1/10 of 1.0 % or less. These systems are already used as operational-type magnetic amplifiers,[13] sometimes in conjunction with a logarithmic network, which is made up of precision resistors and rectifiers.[19] They may also be used as an integrator by introducing a capacitor or a mutual inductance into the degeneration network.[9]

Design Considerations and Experimental Results

SPECIFIC PROBLEM

The high sensitivity of modern forms of permanent-magnet moving-coil-type ink recorders permits in many cases direct operation of these recording milliammeters without the use of amplifiers. There are two basic sensitivities, 0 ma (milliampere) . . . 1.0 ma (resistance, 1,400 ohms), and 0 . . . 5 ma (resistance, 70 ohms). Either type completes its swing across the $4^1/_2$-inch active chart width in about 1/2 second with critical-damping conditions.[20]

In order to extend the range of application of such ink-recording milliammeters, a simple and reliable d-c amplifier is required. Actually, with regard to universal application of a 1.0-ma instrument for recording of direct currents in the range of from 0.1 ma to 1,000 ma, and direct voltages in the range of from 0.5 volt to 1,000 volts, the specific problem consisted in designing a high-precision magnetic amplifier to be operated from the generally available 115-volt 60-cps (cycles per second) line voltage. Since the actual range of this instrument becomes 0. . . ±0.5 ma, when the zero-shift lever in the bottom of its case is shifted to center zero, it is possible to provide also the current ranges of ±0.05 ma to ±500 ma, and the voltage ranges of ±0.25 volt to ±500 volts.

This problem involves, of course, the basic requirement that the complete arrangement will measure currents and voltages within the aforementioned ranges without destroying or changing appreciably what it is trying to measure. Thus, for the maximum value of all current ranges, actual voltage drop across the input terminals should not exceed 0.5 volt or ±0.25 volt, respectively; and, for all voltage ranges, a resistance of 10,000 ohms per volt should be provided.

DIMENSIONS OF COMPONENTS

The afore-mentioned specific problem has been solved by applying the voltage-balance differential-feedback-type self-balancing magnetic amplifier in conjunction with an auxiliary resistance network and a range-selecting switch, as shown in Fig. 7. Table I presents the effective resistance values of the 12 positions of the selecting switch. The ratio of conversion of the current and voltage ranges (4 or 5 respectively) has been selected in such a way that the value to be measured and recorded may be read directly on the linear 50-division scale of the ink recorder.

When measuring currents, the entire current I_X to be measured passes through the selected part of the shunt resistor R_{CR} (R_{0-1} to R_{0-7}, respectively). The actual voltage drop $E_S = 0. . .0.5$ volt across this part of the shunt resistor is balanced by the voltage drop $E_K = I_{LR}R_K = 0. . .0.5$ volt across the negative feedback resistor $R_K = 500$ ohms. When measuring voltages, the current flowing through R_{CR} and the selected part of R_{VR} is proportional to the voltage E_X to be measured. In this case, the actual voltage drop $E_S = E_X R_{CR}/R_{CR}+R_{VR})$ is balanced by the opposing voltage drop E_K. Since the resistance of the potential divider is 10,000 ohms per volt, no excessive current will be drawn from the two points of the circuit under test, with which the input terminals are connected.

The following list gives core and winding data for saturable-reactor elements and power-supply transformer

Table I. Effective Resistance Values Corresponding to the 12 Positions of the Selecting Switch

Switch Position	Effective Resistance, Ohms	Current Range, Ma	Voltage Range, Volts
1	$R_{0-1} = 0.5$	0–1,000	
2	$R_{0-2} = 2.5$	0–200	
3	$R_{0-3} = 10$	0–50	
4	$R_{0-4} = 50$	0–10	
5	$R_{0-5} = 250$	0–2	
6	$R_{0-6} = 1,000$	0–0.5	
7	$R_{0-7} = 5,000$	0–0.1	0–0.5
8	$R_{0-8} = 20,000$		0–2
9	$R_{0-9} = 100,000$		0–10
10	$R_{0-10} = 500,000$		0–50
11	$R_{0-11} = 2,000,000$		0–200
12	$R_{0-12} = 10,000,000$		0–1,000

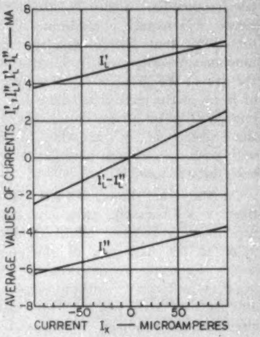


Fig. 8. Average values of output currents I_L' and I_L'', and $I_L'-I_L''$ as a function of the input current I_X to be measured

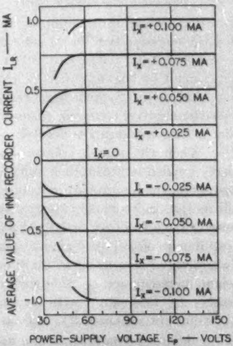


Fig. 9. Average value of resultant output current I_{LR} as a function of power-supply voltage E_P with the current I_X to be measured as a parameter

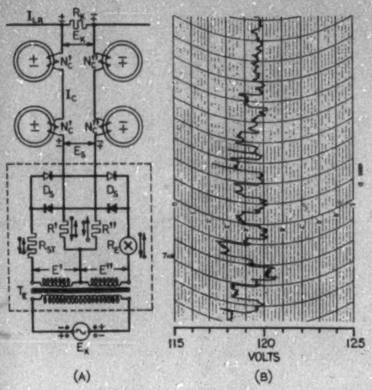


Fig. 10. Application of a voltage-sensitive a-c differential-type network

A—With built-in ring-modulator circuit for recording the fluctuations of the 60-cps power-supply voltage
B—By means of a standard-type moving-coil ink recorder

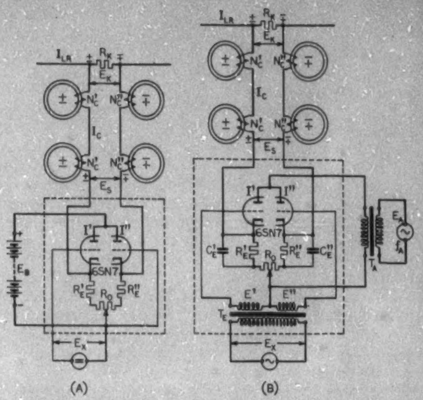


Fig. 11. Application of cathode follower in input circuit of voltage-controlled self-balancing magnetic amplifier of differential-feedback type

A—Using d-c plate supply for measurement of direct voltages E_X
B—Using a-c plate supply for measurement of alternating voltages E_X

T_P (Fig. 1) both for the supply frequency of 60 cps.

Saturable-reactor elements

- Core material: 1/2×0.002-inch superm-alloy tape.
- Core size: inside diameter, $1^3/_8$ inches; outside diameter, $1^3/_4$ inches.
- Load winding N_K: 4,000 turns, no. 34 wire, 220 ohms.
- Bias winding N_B: 2,000 turns, no. 34 wire, 130 ohms.
- Feedback windings N_F' and N_F'': 2× 2,000 turns, no. 34 wire, each 100 ohms.
- Control winding, N_C: 2,000 turns, no. 34 wire, 130 ohms.
- Feedback winding, N_D: 500 turns, no. 34 wire, 35 ohms.

Power-supply transformer

- Core material: 1/2×0.002-inch orthonol tape.
- Core size: inside diameter, $1^1/_4$ inches; outside diameter, 2.0 inches.
- Primary winding: 4,000 turns, no. 33 wire, 230 ohms.
- Secondary windings E_P' and E_P'': 2× 2,000 turns, no. 33 wire, each 100 ohms.
- Secondary winding E_B: 500 turns, no. 33 wire, 32 ohms.

Other pertinent data of the resistance network of Fig. 7 were as follows: $R_M'=R_M''=4{,}000$ ohms; $R_D'=R_D''=100$ ohms; $C_R'=C_R''=4$ microfarads; $R_D=60$ to 80 kilohms. Diffused-junction germanium rectifiers, GE (General Electric Company) type *1N93*, were used, but silicon junction diodes may be used equally well.

Zero adjustment ($I_{LR}=0$ with $E_S=0$) is conveniently achieved by using a variable high-resistance shunt, about 60 to 100 kilohms, across the series-connected bias windings N_B' or N_B''. In this way, it is possible to reduce the magnitude of bias current I_B either in N_B' or in N_B'' in accordance with the actual requirements for perfect symmetry of the push-pull circuit. The regeneration-control resistor R_D will be adjusted so that, for the entire working range ($E_S=0\ldots0.5$ volt), the control-winding current is substantially reduced to zero; $I_C=0\ldots\pm1$ to 2 μa (microamperes).

Results of Measurements

In push-pull self-balancing magnetic amplifiers, the stability may be characterized by the voltage-current relationship which presents the average value of the output current I_{LR} as a function of the supply voltage E_P. When measuring such characteristics on the amplifier (Table I), the magnitude of E_P has been changed within excessively wide limits, 30 to 150 volts, to illustrate the influence of very large changes in supply voltage, which will never occur in practical conditions.

Fig. 8 shows the average values of output currents I_L', I_L'', and $I_L'-I_L''$ as a function of input current $I_X=0\ldots\pm100\mu$a ($E_X=0\ldots0.5$ volt), corresponding to selecting-switch position 7. Fig. 9 presents $I_{LR}=f(E_P)$ with I_X as a parameter. In the entire input-current range (Fig. 8) and also in the supply-voltage range from $E_P=70$ to 150 volts (Fig. 9), the average value of control-winding current I_C varies within the limits of from 0 to ±0.5 to 1.5 μa. The linearity error of the characteristic $I_{LR}=f(I_X)$ is in the order of 1/10 of 1.0%.

The voltage-current characteristics of Fig. 9 are very similar to those of an ordinary saturable-reactor push-pull circuit. The two separately biased saturable-reactor systems retain their inherent negative feedback properties, and the resultant output current $I_{LR}=\text{constant}\times(I_L'-I_L'')$ of these systems with differential feedback is substantially independent of very large changes in supply voltage E_P. Correspondingly, the effects of changes in supply frequency and ambient temperature are minimized.

As a result of these unique performance characteristics, the guaranteed accuracy of the moving-coil ink-recorder system, 1.0% of full-scale current, is not appreciably affected by introduction of the amplifier. The response speed and damping of the ink recorder can be varied in the usual way by means of a variable high-resistance shunt across the terminals of the recorder. The response-speed characteristics are very similar to those obtained without the use of the amplifier.[20] With critical-damping conditions, the response time is about 1/2 second. When shifting the zero of the recorder from the left to center ($I_{LR}=0\ldots\pm0.5$ ma), the response-speed characteristics remain unchanged.

Fig. 10(A) shows a novel form of a voltage-sensitive differential-type network with built-in ring-modulator circuit, which compares the actual value of the nonlinear resistance R_E with the value of standard resistor R_{ST}. The polarity-reversible d-c signal voltage E_S across the mixing resistors R' and R'' is proportional to actual deviation of the alternating voltage E_X from its rated value at which the differential network is balanced ($R_E=R_{ST}$). Since E_S is balanced by the voltage drop E_K across resistor R_K, the average value of the resultant load current I_{LR} flowing through the ink recorder is nearly proportional to the actual deviation of E_X from its rated value. Fig. 10(B) shows the fluctuations of the 60-cps power-supply voltage which were recorded by means of the voltage-sensitive circuit of Fig. 10(A), using the following data of components:

R_E, GE lamp, 120 volts, 3 watts; R_{ST}, wire-wound resistor, 2,930 ohms.
$R'=R''$, wire-wound resistors, each 2,000 ohms (mixing resistors).
D_S, germanium diodes, GE type *1N93*; T_E, step-down transformer, 10 volt-amperes.
$E_X=115\ldots120\ldots125$ volts (10 volts corresponding to 50 scale divisions).
$E'=E''=57.5\ldots60\ldots62.5$ volts; $E_S=0\ldots\pm0.25$ volt (average value).

Finally, with regard to the internal resistance of the d-c signal source of voltage-controlled self-balancing magnetic amplifiers, note that this resistance should not exceed 5,000 to 10,000 ohms. Otherwise, the polarity-reversible compensating-current component, equation 6, providing the negative-feedback ampere-turns I_KN_C' and I_KN_C'' will be too small for securing the very high stability of the amplifier.

Fig. 11(A) illustrates the possibility of employing a conventional cathode-follower circuit with d-c plate supply in conjunction with a voltage-balance arrangement as in Figs. 1 and 2. In this case, the signal-source resistance of the magnetic amplifier itself is $R_E'+R_O+R_E''=6{,}000$ ohms, and the input resistance of the cathode-follower circuit, as seen from the terminals of the signal source E_X, is in the order of 1,000 megohms or higher (measurement of *pH* values, etc.). The modified circuit of Fig. 11(B) which uses a-c plate supply offers the possibility of applying a phase-reversible a-c signal voltage E_X which is synchronous with the plate voltage E_A (frequency f_A). In this way, the range of application may be extended into the realm of those a-c measurements which are based upon the use of phase-sensitive rectifiers (vector meter).

An important advantage of the arrangements of Fig. 11 is the fact that the average value of control-winding current $I_C=\pm1$ to 2 μa is very small in comparison with the magnitude of the plate currents I' and I'' (about 1 to 2 ma). Thus, because no appreciable current I_C is drawn from the cathode resistors R_E' and R_E'', the operation of the cathode-follower circuit remains undisturbed. Furthermore, because the actual output power derived from R_E' and R_E'' is extremely small, about 10^{-9} watts, the cathode follower may be operated with considerably reduced values of plate voltage and plate currents. This possibility is advantageous with regard to the life of the vacuum tube.

References

1. Arrangement for Measuring Small Direct-current Quantities (Einrichtung zum Messen von Kleinen Gleichstromgrössen), **W. A. Geyger.** *German Patent no. 702,082*, Jan. 30, 1941.

2. Fundamentals of Magnetic Amplifiers for the Techniques of Measurement and Control (Grundlagen der magnetischen Verstärker für die Mess- und Regeltechnik), **W. A. Geyger.** *Wissenschaftliche Veröffentlichungen aus den Siemens-Werken*, Berlin-Siemensstadt, Germany, vol. 19, Nov. 1940, pp. 226–69.

3. A High-sensitivity Magnetic Amplifier of the Balance Detector Type for the Techniques of Measurement and Control (Ein hochempfindlicher magnetischer Verstärker für die Mess- und Regeltechnik), **W. A. Geyger.** *Ibid.*, vol. 21, 1943, pp. 47–57.

4. Magnetic Amplifiers of the Balance Detector Type—Their Basic Principles, Characteristics, and Applications, **W. A. Geyger.** *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1707–20.

5. *Discussion* by **C. S. Hudson** on the papers by Mr. Milnes and Messrs. Gale and Atkinson. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 96, pt. I, Dec. 1948, p. 115.

6. Magnetic Amplifiers, **A. G. Milnes.** *Ibid.*, no. 99, May 1949, Fig. 11, pp. 89–98.

7. Improvements in and Relating to Magnetic Amplifiers, **C. S. Hudson.** *British Patent no. 630,536*, Oct. 14, 1949.

8. Improvements in Saturable Core Reactance Devices, **F. H. Belsey, J. S. Wilson.** *British Patent no. 659,762*, Oct. 24, 1951.

9. Magnetic Amplifiers of the Self-Balancing Potentiometer Type, **W. A. Geyger.** *AIEE Transactions*, vol. 71, pt. I, Jan. 1952, pp. 383–95.

10. Self-balancing Magnetic Amplifier, **W. A. Geyger.** *U. S. Patent 2,700,130*, Jan. 18, 1955.

11. Magnetic Amplifiers (book), **H. F. Storm.** John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 213 and 225.

12. Measurements on High-Speed Magnetic Servo Amplifiers with 2-Phase Motor Load, **W. A. Geyger.** *AIEE Transactions*, vol. 75, pt. I, May 1956, pp. 242–48.

13. An Operational-Type Magnetic Amplifier for Air-borne Servo-Control Systems, **R. M. Hubbard.** *Ibid.*, vol. 73, pt. I, Nov. 1954, pp. 425–30.

14. On the Mechanics of Magnetic Amplifier Operation, **R. A. Ramey.** *Ibid.*, vol. 70, pt. II, 1951, pp. 1214–23.

15. Fast Response with Magnetic Amplifiers, **D. G. Scorgie.** *Ibid.*, vol. 72, pt. I, Jan. 1954, pp. 741–49.

16. The Operation of the Self-Balancing Magnetic Amplifier, **Albert D. Krall, Edward T. Hooper.** *Ibid.*, vol. 75, pt. I, Mar. 1956, pp. 79–84.

17. A Novel Principle of Direct-Current Amplification (Ein neuartiges Prinzip stabiler Gleichstrom-Verstärkung), **R. Eberhardt, G. Nüsslein, H. Rupp.** *Archiv für Elektrotechnik*, Berlin-Charlottenburg, Germany, vol. 35, no. 8, 1941, pp. 477–89; no. 9, 1941, pp. 533–49.

18. Analog Methods in Computation and Simulation (book), **W. W. Soroka.** McGraw-Hill Book Company, Inc., New York, N. Y., 1954, pp. 75–79, and pp. 82–86.

10. An Analogue Computer Technique Using Magnetic Amplifiers, **B. E. Davis, I. H. Swift.** *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 635–40.

20. Esterline-Angus Recording Permanent-Magnet, Moving-Coil Type Direct-Current Milliammeters. *Catalog Section 42–56*, The Esterline-Angus Company, Inc., Indianapolis, Ind., 1956, pp. 1–8, Figs. 5 and 8.

21. Separately Controlled Rectifiers for Measurement Purposes, Circuit Arrangements with Vacuum Tubes (Fremdgesteuerte Mess-Gleichrichter, Schaltungsanordnungen mit Elektronenröhren), **W. A. Geyger.** *Archiv für Technisches Messen*, Munich, Germany, Section Z 53-1, no. 162, July 1949.

Characteristics and Applications of the Iatron Storage Tube

DEAN W. DAVIS
NONMEMBER AIEE

THE OPERATION of many types of storage tubes depends upon intensity modulation of an electron beam by an electrostatic charge on an insulator layer.[3-7] As early as 1927, Dr. Farnsworth suggested the method as a means to increase brightness in cathode-ray tubes for television.

Iatron* storage tubes have been in development at Farnsworth Electronics Company since 1949, and have been in experimental use since 1953. The properties of image storage and extremely high brightness at low voltage make the tube very attractive for radar indicators, oscilloscopes, and other uses.

This paper describes its operation with emphasis on those unusual characteristics which have no counterpart in ordinary cathode-ray tubes, and suggests modifications of an oscilloscope to use the tube. It is hoped that the applications engineer will find the answers to some of the questions which may arise when operating this storage tube for the first time.

Description

The Iatron is a storage cathode-ray tube in which the display can be written, stored, and viewed continuously. While it is available in a variety of bulb sizes and types of writing guns, the characteristics presented here are particularly applicable to the 5-inch *Ia10P20-25* tube, shown in Fig. 1, which has an electrostatic writing gun. Continuous viewing is provided by a divergent electron beam, called the flooding beam, which expands until its spot covers the display area of the tube. Elemental areas of the large flood spot are modulated by means of a cathode-ray beam of the type used in conventional cathode-ray tubes.

The following explanation of the method of modulating the flood spot refers to Fig. 2. The flooding gun is located on the tube axis in the rear wall of the bulb. The control grid of the flooding beam is an insulator layer which covers the front surface of a fine-mesh metal screen, insulator screen, and is located in the path of the beam at the distance where the beam has expanded to its maximum size. The beam passes through the metal screen and impinges upon the aluminized phosphor on the inner face of the tube. An aquadag coating on the wall of the bulb and a metal collector screen serve to collimate the flooding beam. With proper collimation the paths of the electrons in the expanded flooding beam are made parallel to each other and perpendicular to the plane of the insulator.

The insulator screen which supports the insulator layer serves the same purpose as the screen grid of a tetrode and is operated on a fixed voltage of about +10 volts. Points on the insulator surface, however, can assume potentials which are quite different from each other and from the voltage on the insulator screen. These potentials are established on the insulator surface by charging it with the writing beam and with the flooding beam. Over the insulator control range, the number of flooding electrons which can pass through the screen at any point is roughly proportional to the insulator surface potential at that point.

ERASING

Erasing pulses are essential to the operation of the Iatron. These are low-voltage pulses applied to the insulator screen which cause the display to fade to cutoff. Without them, the entire display area would increase eventually to maximum brightness. The increase in brightness results when positive ions, which are generated by collision between flooding electrons and residual gas molecules, are drawn to the insulator surface charging it positively. The ion density over the insulator surface is uniform and therefore the display area brightens everywhere at the same rate.

Further increase of insulator potential ceases when the insulator has become charged to about +2 volts with respect to the flooding-gun cathode. At this potential, some flooding electrons have enough energy to strike the insulator. Striking with low energy, these electrons stick and by virtue of their negative charge prevent additional voltage increase. An equilibrium insulator potential is thereby established for which the positive ions are balanced by the electrons landing on the insulator.

When the insulator suddenly becomes more positive, as by applying a voltage pulse to the support screen, flooding current strikes the insulator, quickly charging the surface down to equilibrium again. At the end of the pulse, the support screen is returned to its original potential, but the insulator is now more negative than it was before because of the charge acquired during the pulse. Positive insulator charges written by the writing beam are erased in the same way.

INSULATOR CONTROL CHARACTERISTIC

Fig. 3 is a curve showing average phosphor flooding current over the display area as a function of insulator surface potential. The points on the curve were obtained, using +10-volts bias on the support screen, by first allowing ions to charge the entire insulator surface to equilibrium. At equilibrium, the insulator is charged to its most positive potential and phosphor current is a maximum. This current corresponds to zero insulator volts on the curve. The insulator surface was then charged negatively 0.5 volt by applying a +0.5-volt pulse to the insulator support screen. The phosphor current was measured at the end of the pulse, and the process was repeated increasing the amplitude of the pulse at each step until phosphor current reached cutoff. Thus it was determined that pulses of 3.2 volts amplitude are necessary to erase the tube to cutoff. The assumptions were made that the insulator surface is at zero volts at equilibrium, and that the insulator is charged negatively to the amplitude of the erase pulse at each step. Although, as a matter of academic interest, the insulator surface potential

---

* Iatron is a trade-mark of the Farnsworth Electronics Company, a division of International Telephone and Telegraph Corporation.

Paper 57-7, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N.Y., January 21-25, 1957. Manuscript submitted June 15, 1956; made available for printing October 29; 1956.

DEAN W. DAVIS is with the Farnsworth Electronics Company, Division of International Telephone and Telegraph Corporation, Fort Wayne, Ind.

Dr. P. T. Farnsworth has made many notable contributions to the storage-tube art[1,2] and has personally directed the Iatron research and development work undertaken in 1949. Others who have contributed to the development of the Iatron include a majority of the Farnsworth Research Department. Current Iatron development has been supported primarily by the U. S. Navy, Bureau of Ships.

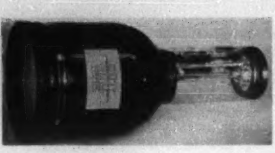


Fig. 1. 5-inch electrostatic Iatron IA10P20-25

at equilibrium was actually about 2 volts, the first assumption is valid in the sense that the insulator does not draw flooding current over the control range from equilibrium to cutoff.

Control of Viewing Time

It is clear that the insulator cannot be charged to a potential more negative than the absolute amplitude of the erasing pulse, because it charges to an equilibrium potential which is reached when the voltage difference between the flooding-gun cathode and insulator is just sufficient for flooding electrons to strike the insulator. On the other hand, it can be prevented from charging to this equilibrium if the pulse duration is shorter than the charging time required. The erasing pulses used to obtain the curve of Fig. 3 were applied for a sufficient length of time to charge the insulator to equilibrium.

The average time required to charge the insulator from maximum brightness to cutoff is about 3 milliseconds, and a display can be made to persist for many seconds if pulses which are narrower than 3 milliseconds are applied continuously at a suitable repetition rate. On each successive pulse, the insulator will be charged negatively a small amount and will eventually reach cutoff when the product of pulse width times the number of consecutive pulses is equal to 3 milliseconds.

Pulses having a repetition frequency of f pulses/sec (per second), which are t sec wide and 3.2 volts in amplitude, will therefore erase signals from maximum brightness to cutoff in the approximate time $T=0.003/f\times t$ sec, for large values of f. (The erasing-pulse amplitude required to erase to cutoff is increased slightly if the erasing-pulse frequency is reduced to very low values, since the insulator voltage shift by ions during the interval between pulses then becomes significant.) Hence, the viewing time of the display can be varied by controlling either the frequency or the width of the erasing pulses. The proper choice of pulse repetition is influenced by flicker and contrast as well as viewing time.

Flicker

During an erasing pulse, the entire display area is illuminated by flooding current which passes through the insulator screen (see Figs. 5 through 7). The resulting flashes of light produce flicker at low pulse frequencies, or merge to a constant background brightness at a pulse frequency of about 45 pulses/sec. However, very narrow pulses such as would be used to obtain a viewing time of the order of 1.0 minute are hardly detectable in the display even in the flicker frequency range.

Contrast

Loss of contrast is caused by background brightness. The average background contributed by erasing pulses is $B=0.003/T(100)=(100\ f\times t)$per cent of maximum display brightness.

Loss of contrast would be most severe using the erasing conditions required for minimum viewing time. However, this is an unrealistic condition for the Iatron as it defeats the very purpose of storage tubes. Nevertheless, it is interesting to estimate the maximum background brightness which might be encountered.

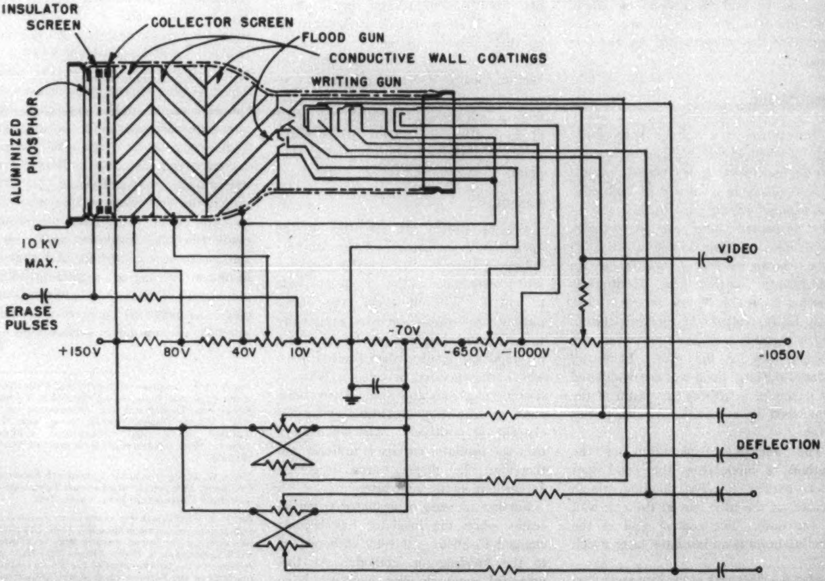


Fig. 2. Iatron model IA10P20-25, schematic

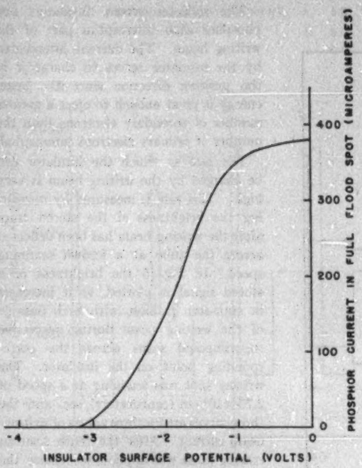


Fig. 3. Static insulator control characteristic

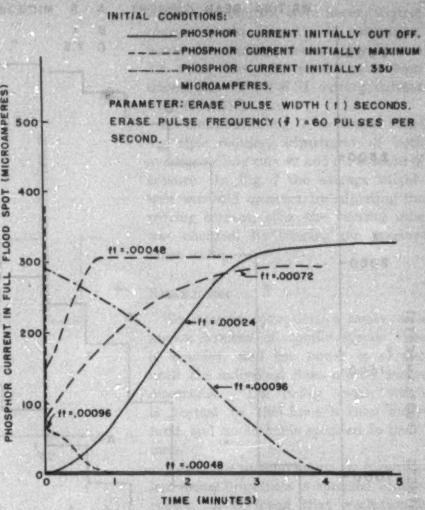


Fig. 4. Change of phosphor current caused by positive ions and erasing pulses

Since the minimum erasing time is 3 milliseconds and the repetition rate over which persistence is limited by the eyes of the observer is 45 pulses/sec, these values may be taken as the pulse-width and frequency which will determine minimum viewing time. Under these conditions, the display area would be viewed at maximum brightness $45 \times 0.003 = 0.135$ sec/sec of total viewing time. The average background brightness would, therefore, be about 13.5 per cent of the maximum signal brightness.

However, by the same reasoning, B would be only 0.3 per cent for a viewing time of 1.0 sec. At 45 pulses/sec, the pulse width for this condition would be about 67 microseconds.

MAXIMUM VIEWING TIME

Viewing time may be limited by positive ions or by insulator leakage. Leakage is negligible at normal tube operating temperatures, and most Iatrons will store written charges for several hours if all tube voltages are removed, to avoid generation of ions, after writing.

The number of ions generated is a function of several things such as flooding current density, length of the electron paths, pressure of residual gases in the tube, etc. The time required in an average tube for ions to charge the insulator from cutoff to 50 per cent of maximum brightness is about 30 sec. The measurement is significant in that there must be a minimum product of erasing frequency and pulse width, $f \times t$, which will prevent ion charge integration on the insulator and thereby maintain nearly constant insulator potential in the interval between pulses.

Fig. 4 shows curves of phosphor current versus time for a typical tube using small values of $f \times t$. The ordinate is proportional to brightness since the flooding-spot size is constant and the power input to the phosphor is well below the level at which saturation occurs. For the condition that the insulator is initially at cutoff, an erasing product $f \times t = 0.00048$ maintains cutoff. If the insulator is initially at equilibrium however (phosphor current initially maximum), an erasing product $f \times t = 0.00048$ will not erase the tube but will allow the insulator to charge to an intermediate potential corresponding to 310 μa (microamperes) of phosphor current. Thus it is found that two stable potentials, one corresponding to cutoff and the other corresponding to 310 μa of phosphor current, are possible. They are maintained in each case because the insulator charge received from the flooding electrons during one erasing pulse is equal to the ion charge acquired during the interval between pulses.

A necessary condition for continuous operation is that signals written by the writing beam to any brightness will eventually be erased to cutoff. Lacking sufficient erasure, the display area would soon be saturated. As shown in Fig. 4, if a minimum erasing product $f \times t = 0.00096$ is used, a signal which has been written to maximum brightness will assuredly be erased.

Two curves are shown for which $f \times t = 0.00096$. In the first, the pulse was not applied until the insulator had been charged to equilibrium and allowed to remain there for a considerable length of time. By suddenly applying the pulses, phosphor current drops sharply and the tube erases to cutoff. In the second curve, erasing pulses were not applied suddenly but were operated continually while the tube was written to the initial phosphor current of 330 μa. This curve is typical of viewing-time curves obtained with continuously operating erasing pulses, the viewing time being less for larger values of $f \times t$.

The cause of the rapid decrease in brightness which is observed when an erasing pulse is suddenly applied, under the initial condition that the phosphor current is a maximum, is not fully understood. However, it is probable that the distribution of charges on the insulator on the front surface as well as on the sides of the mesh holes of the insulator screen is affected by initial conditions.

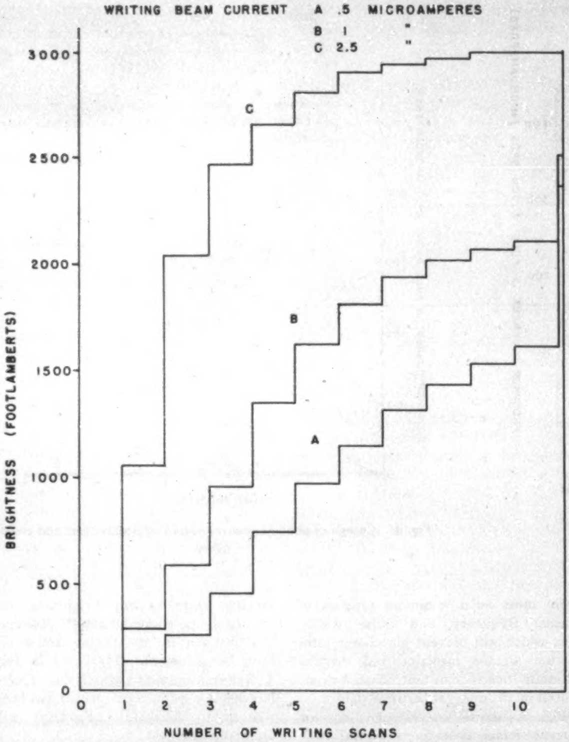


Fig. 5. Brightness of a point written upon for ten successive writing scans and finally erased to cutoff; writing-spot scanning speed 2.75×10^4 cm/sec

Depending upon the distribution of charges, the electric fields in the vicinity of the meshes will be modified, deflecting the flooding electrons and ions to selective minute areas of the insulator surface and, consequently, altering the erasing characteristic.

It is interesting to note that if the insulator is initially at cutoff, and a value of $f \times t$ is used so that $0.00048 < f \times t < 0.00096$, insulator areas where a large writing charge has been deposited will charge to a definite brightness level, while weak signals will be erased to cutoff. (A small area of positive charge will not persist indefinitely however, since the escape of secondary electrons from the area is inhibited by the electric field of the less-positive surrounding insulator surface, thus causing more electrons to stick, charging the area negatively. The bright area shrinks in size and disappears.) This characteristic may be used to achieve extended viewing time but with concurrent loss of half tones. Extension of viewing time by this method is even greater if still narrower erasing pulses of relatively high amplitude are used. The viewing time obtainable in this way is upwards of 30 sec after which time the written areas decay rather rapidly to cutoff.

WRITING CHARACTERISTIC

With the tube operating and with proper erasing pulses being applied, the tube will be at cutoff and in readiness to be written upon. The 1,000-volt writing beam may be scanned over the insulator in any desired pattern, and video signals are applied to the control grid to modulate the beam.

The collector screen, insulator, and phosphor each intercept a part of the writing beam. The current intercepted by the insulator serves to charge it in the positive direction since the beam energy is great enough to eject a greater number of secondary electrons than the number of primary electrons intercepted.

The rate at which the insulator can be charged by the writing beam is very high. This rate is measured by measuring the brightness of the stored trace after the writing beam has been deflected across the tube at a known scanning speed. In Fig. 5 the brightness of a stored signal is plotted, as it increases in stair-step fashion, with each passage of the writing beam during successive superimposed scans across the corresponding point on the insulator. The writing spot was scanning at a speed of 2.75×10^4 cm (centimeters)/sec, and the three curves are for three values of writing-beam current. After the tenth scan an erasing pulse was applied to restore the insulator to cutoff. The bright flash which accompanies the erase pulse is shown also in Fig. 5. It can be seen that the flash is never brighter than the brightness associated with equilibrium insulator potential, and has a relationship to the brightness of the signal which is being erased.

The writing currents used to obtain the data shown in Fig. 5 were necessarily very low because of the relatively low scanning speed available for these tests. However, the writing-beam current can be as high as 150 μa at zero grid bias and will write to maximum brightness in a single-line scan at a scanning speed of about 10^5 cm/sec.

At high brightness, the brightness increase per scan grows smaller. At the high-brightness limit saturation is reached, and further writing results only in a charge spreading on the insulator surface. Since an electron beam does not possess finite size, a few electrons will be found even at radii considerably larger than the dimensions of the spot defined by the usual methods. These fringe electrons continue to integrate on the insulator area surrounding the core of the spot after it has saturated.

The obvious implication is a serious one, that areas repeatedly written upon in a display will tend to bloom unless some form of insulator writing-charge limiting is used. In a plan position indicator radar display, for example, an equalizing signal derived from the range-deflection voltage may be added to the video signal to prevent blooming at the center. To emphasize weak targets

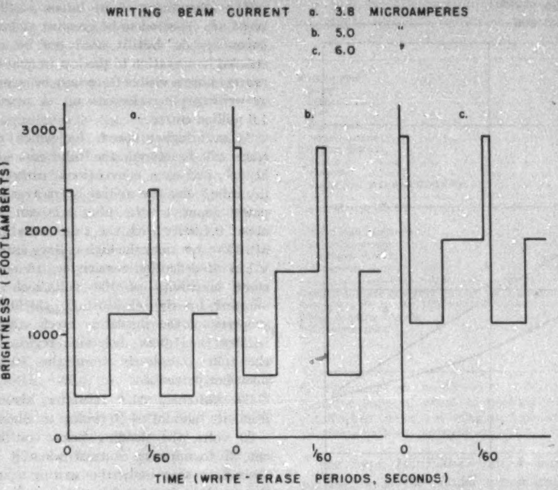


Fig. 6. Stable brightness levels of a point alternately written and erased upon; erase-pulse width 1,350 microseconds

in a display, video clipping will prevent blooming on strong signals. Proper choice of the erasing-pulse frequency is, of course, extremely important in this matter since writing charge can be erased before it has had the opportunity to integrate beyond a desired level. An erasing frequency equal to the deflection frequency is the best choice in many cases.

In almost any continuous display the erasing-pulse frequency and pulse width are preset, i.e., after adjustments have been made to achieve the desired viewing time no further adjustments will be required during operation. Fig. 6 illustrates typical continuous operation of an Iatron using an erasing frequency equal to the writing-beam scanning frequency, which in this case is 60 cps (cycles per second). An erasing-pulse width of 0.00135 sec was used, and the viewing time was consequently about 37 milliseconds. The way in which the average brightness varies with writing current is apparent from the three conditions shown. The two brightness levels in each graph correspond to insulator potentials immediately after writing and immediately after erasing. These are stable potentials which will be repeated until either writing current or erasing-pulse width is changed. Each erasing pulse charges the insulator downward in potential by an amount equal to the potential increase by writing during one scan. The ability of the insulator to assume stable potentials follows from its nonlinear charging characteristic wherein the charging rate decreases toward higher brightness levels; Fig. 5. If the characteristic were linear, writing charges would integrate to saturation. Since the erasing-pulse amplitude was 3.2 volts, the tube would have been erased just to cutoff if writing current were reduced to zero.

As indicated in Fig. 7, to change viewing time requires adjustment of both average writing current and the amount of erasure. In Fig. 7 the average brightness was held constant by adjusting the writing current, after the viewing time was changed, by altering the erasing-pulse width.

RESOLUTION

To measure resolution, a raster of a known number of equally spaced lines is scanned, and the raster is shrunk until the individual lines are no longer discernible. The raster width which is normal to the lines is then measured, and resolution is specified in lines/inch.

Resolution measurements are meaningless unless brightness is specified concurrently. It is found that resolution is approximately the same at a given brightness regardless of how many scans or what writing-beam current were required to write to that brightness. While the resolution of a stored image at low brightness is nearly equal to the resolution of the writing beam itself, at high brightness it approaches a minimum of about 35 lines/inch; see Fig. 8.

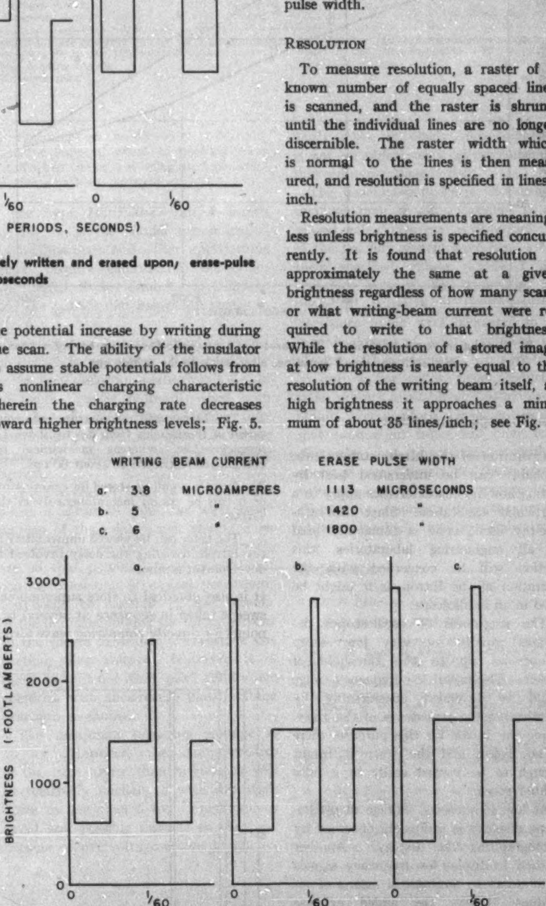


Fig. 7. Stable brightness levels of a point alternately written and erased upon; average brightness is constant

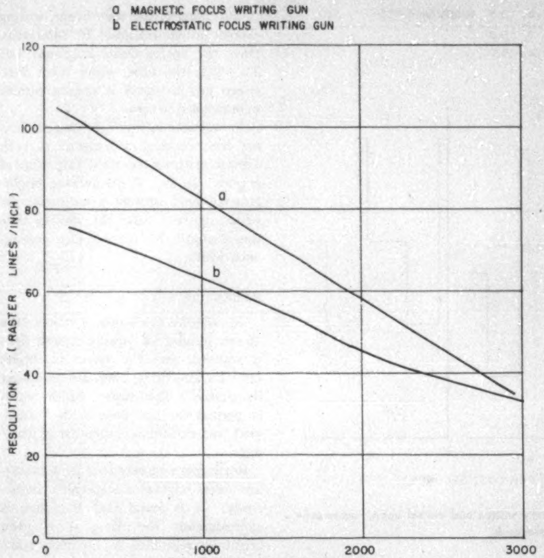


Fig. 8. Resolution versus brightness. The curve for a magnetic focus gun having a smaller writing-spot size shows limitations attributable to the writing gun

Oscilloscope Application

Operation of the Iatron storage tube probably can be understood best by noting how its characteristics apply to a particular application. Since the cathode-ray oscilloscope is commonly used in all engineering laboratories, this section will be concerned with the operation of the Iatron as it might be used in an oscilloscope.

The usefulness of oscilloscopes decreases rapidly for very low sweep frequencies up to the threshold of flicker. The useful low-frequency range could be extended considerably by lengthening the persistence of the trace. Using the Iatron for this purpose, there is no flicker and the trace is bright enough to be viewed easily in a fully lighted room.

At low frequencies, storage in oscilloscope displays is ordinarily obtained by photographing the display. Another method to display low-frequency signals is to resort to mechanical means of recording. Besides the added expense and inconvenience of these methods, they also have limitations which are overcome by using the Iatron:

1. The Iatron will record transients composed of frequencies from d-c to above 1.0 megacycle/sec, whereas mechanical recorders are restricted to about 60 cps.
2. Any trace can be stored for examination or it can be erased instantaneously in the Iatron.
3. The trace can be viewed immediately in the Iatron, avoiding the delay involved in development of film.

It is also practical to store superimposed sweeps taken in sequence at several test points for directly comparing wave forms.

The advantages of an Iatron oscilloscope are expected to be greatest at low-sweep speeds, but it need not be restricted in operation to the low-frequency ranges, since a visible trace may be stored at writing-spot velocities up to nearly 1.0 million cm/sec.

At still higher speeds for which no trace will be stored, the tube can still be operated as a conventional cathode-ray tube, since the writing-beam average power input to the phosphor can be about 0.4 watt with the tube operating at 10 kv; nor does the high voltage entail a loss of deflection sensitivity. A constant sensitivity of 100 volts/inch is afforded by the electrostatic shielding property of the insulator screen which isolates the 1.0-kv deflection region of the tube completely from the 10-kv phosphor potential.

By switching the insulator screen from its normal +10 volts to about −20 volts, the flooding beam can be cut off to improve contrast when it is desired to view only the writing-beam trace and not its stored image.

In normal operation, erasing pulses will keep the insulator erased to cutoff in areas where no trace is being written, and will prevent writing charges in the trace from integrating to the extent of charge spreading. For the usual repetitive-signal mode of oscilloscope operation, an erasing-pulse amplitude control and erasing-pulse width control should be accessible on the front panel to make initial adjustments of cutoff and viewing time. An intensity control is necessary to adjust writing-beam current to compensate for changes in sweep speed and waveform of the signal.

At sweep frequencies of over 45 cps the erasing pulse can be triggered by the sweep. This is in keeping with the discussion of writing-charge limiting in which it was pointed out that maximum charge stability of areas repeatedly

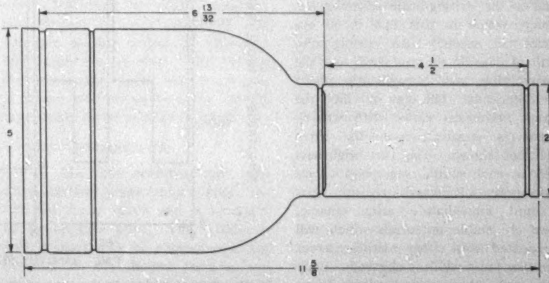


Fig. 9. Outside dimensions of the 5-inch electrostatic Iatron IA10P20-25

written upon exists when writing and erasing frequencies are equal. At lower sweep frequencies, flicker and blooming would be avoided if a constant erasing-pulse frequency of 45 cps or higher were used. A convenient and satisfactory frequency is 60 cps.

To display transients, maximum writing speed and storage time is desirable. A switch might be provided which, after writing, could be used to cut off the writing beam and erasing pulses simultaneously, thus avoiding overwriting of the transient trace and at the same time preventing its erasure. At extremely slow sweep speeds it is desirable to turn off the erase before the start of the writing trace to avoid any erasure before one sweep is completed. This suggests a manual on-off erase switch. Also, an instant-erase button would probably be useful to restore the insulator quickly to cutoff after operating with the erasing pulse off.

The oscilloscope should be equipped with a *z*-axis gate to assure that the undeflected writing spot is cut off, since the undeflected spot would cause insulator charge spreading from that spot over an appreciable area of the screen, and at very high current the insulator might even be damaged.

The controls described are the extent of the added complexity necessary to operate an oscilloscope adapted to the Iatron, and the additional circuitry needed to operate the flooding system is equivalent to adding one tube. An erasing-pulse generator which can perform the suggested functions could be a slave multivibrator.

Summarizing, the following controls are recommended for full utilization of the tube's capabilities:

1. Erasing-pulse gain control to adjust the amplitude of the pulses.
2. Erasing-pulse width control to adjust the duty cycle of the pulses.
3. Instant erase push-button switch to widen the erasing pulses momentarily in order to erase clutter without disturbing other erasing control settings.
4. Erase on-off switch to remove erasing pulses when it is desired to "freeze" a trace for inspection without adjusting erasing controls.
5. Write on-off switch to bias the writing-gun control grid to cutoff to prevent overwriting a "frozen" trace without adjusting the intensity control.
6. Flooding-beam on-off switch to bias the insulator support screen to flooding-beam cutoff when the tube is being used as a conventional cathode-ray tube.

Table I. Operating Voltages

	Volts	Current
Writing gun		
heater	6.3	0.6 amp, a-c or d-c
cathode	−1,000	1,080 μa, max
grid	−1,042 at cutoff	
first anode	−700 at focus	−1.0 μa, max
second anode	40	940 μa, max
Flooding system		
heater	2.5	2.5 amp, a-c or d-c
cathode	0	2.6 ma
anode and first wall electrode	40	0.8 ma, min 1.0 ma, max
second wall electrode	20	0.07 ma, min 0.115 ma, max
third wall electrode	80	0.035 ma
collector screen	150	1.25 ma
insulator screen	+10	
phosphor	+10*	0.38 ma, max

Deflection-plate reference voltage for minimum astigmatism, 0 volts. Deflection sensitivity; 85 volts/inch, plate D_3 to D_4; 100 volts/inch plate, D_1 to D_2. Plates D_1 and D_2 connected to +90 volts draw 36-ma flooding current.

* Kilovolts, max.

The type *Ia10P20-25* Iatron shown in Fig. 1 is the model recommended for oscilloscopes and other applications which require electrostatic deflection. The useful display diameter is 4 inches, and the outside dimensions are shown in Fig. 9.

Some comment on the operating circuit of Fig. 2 is necessary. The resistance in the final deflection-plate circuit should be lower than is ordinarily used with cathode-ray tubes, because when they are driven positive, the plates can draw about 36 microamperes of flooding current.

If the average voltage of the deflection plates is about 40 volts, the least astigmatism of the writing spot results since the second anode and first wall electrode are on that potential. An astigmatism control consisting of a dual potentiometer could be inserted in the bleeder at the points supplying the direct current to the plates to adjust the average deflection plate voltage. However, it is found in practice that good results are achieved with an average plate voltage near zero, as shown.

The maximum phosphor voltage is 10 kv. However, the characteristics of the tube, other than brightness, will be relatively unchanged with operation down to less than 5 kv. Therefore, to avoid any possible damage to the tube because of overvoltage accidents, particularly when extremely high brightness is not an objective, it is recommended that reduced voltage be used.

The flooding-spot size is adjusted by small changes in voltage applied to the second wall electrode after other voltages of the flooding system have been set at their specified values. Table I lists operating voltages and maximum and minimum currents of flooding-system electrodes which were measured to aid in the design of power supplies and bleeder. These measurements were made on only a few tubes, since production tubes are not available at this writing to obtain average data. It is anticipated that production tubes will have fewer electrodes, but those retained will be operated very closely to their present voltages, and tubes will operate interchangeably, requiring only the number of controls which have been indicated.

References

1. *U. S. Patent no. 2,228,338.*

2. *U. S. Patent no. 2,754,449.*

3. STORAGE TUBES AND THEIR BASIC PRINCIPLES (book), **M. Knoll, B. Kazan.** John Wiley & Sons, Inc., New York, N. Y., 1952.

4. CHARACTERISTICS OF A TRANSMISSION CONTROL VIEWING STORAGE TUBE WITH HALFTONE DISPLAY, **M. Knoll, H. O. Hook, R. P. Stone.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 42, Oct. 1954, pp. 1496–1504.

5. DIRECT VIEWING MEMORY TUBE, **S. T. Smith, H. E. Brown.** *Ibid.*, vol. 41, Sept. 1953, pp. 1167–71.

6. THE RECORDING STORAGE TUBE, **R. C. Hergenrother, B. C. Gardner.** *Ibid.*, vol. 39, 1950, pp. 740–47.

7. A MEMORY TUBE, **A. V. Haeff.** *Electronics*, New York, N. Y., vol. 20, Sept. 1947, pp. 80–83.

Where We Are and Where We Are Going in Telephone Transmission

H. R. HUNTLEY
FELLOW AIEE

THE PRINCIPAL business of the telephone industry is to provide circuits and switching arrangements so that people who are physically separated can talk to each other. It also is in the business of transmitting other forms of information (teletypewriter, data, radio programs, television, etc.) from one place to another. The things which affect talking and transmitting other information are called "transmission."[1,2]

Transmission is such an important factor in the telephone business that its basic objectives must be the same as the basic objectives of the business itself, which are the same as those of any other economic enterprise; that is:

1. It must create customers and keep them happy.
2. It must take in enough money to pay its employees adequately, to satisfy its creditors, and to make enough profit so that many people will want to share in its ownership.

From these basic objectives and even from a rudimentary understanding of people and their motivation, it is possible to derive more concrete objectives applying to transmission, namely:

1. Transmission must get more pleasing and attractive to customers continuously.
2. It must encompass an ever-widening field as customers' views and needs change.
3. Its cost must be kept as low as practicable.

Over-all Progress

At the start, it may be well to summarize the progress which has been made in telephone transmission and which seems likely in the reasonably near future.

Under present conditions, the best over-all single measure of ease of talking is the ratio between the acoustic pressure (weighted in accordance with loudness) of the sounds in the listener's ear and the acoustic pressure (similarly weighted) of the sounds at the talker's lips.[3] The acoustic pressure of the sounds reaching the listener's ear is a reasonable measure of how easy it is for him to hear. The acoustic pressure of the sounds which the talker has to put out in order that the listener may hear easily is a measure of how hard he has to work. The ratio of these acoustic pressures is a pretty good over-all measure of the combined efforts of the talker and listener.

This ratio of acoustic pressure on typical calls has improved greatly in the past and it can reasonably be expected to improve further in the future. Table I compares this ratio today with conditions 25 years ago and makes a prediction of what is expected 5 or 10 years hence. The figures for typical calls 5 to 10 years hence are roughly equivalent to talking at a few feet apart in open air, such as across a desk.

There are factors other than acoustic pressure ratio which affect the ease and pleasure of talking, such as sidetone (that part of one's own voice which he hears directly in his own receiver), echo (one's own voice, but delayed because it has traveled out to some point in the circuit and back again), bandwidth, crosstalk (interaction between different circuits), amplitude distortion (usually manifest as making voices sound raspy), and noise.

With modern instrumentalities and methods these impairments can usually be kept small enough to be unimportant, but if any one of them exists, it can be taken into account by appropriate corrections. Noise is probably the most difficult of the impairments to control and constant effort is required to prevent it from becoming significant.

What People Want

If the telephone business is to create customers and keep them happy, it must know not only how people like the transmission they are getting, but also what they would like if there were no technical or economic problems to be overcome.

A wide variety of things are done to answer these questions, including organized programs of asking customers what they think, judgment tests by telephone people, from a wide variety of jobs, serving as guinea pigs, etc. The results of all of these various methods are reasonably consistent among themselves and with field experience and can be summarized briefly as follows:

1. People, especially those who make a large number and variety of calls, are quite critical and will call transmission only "fair" or even "poor" on calls on which it is perfectly possible to talk without repeating at all, or at most, only occasionally.
2. Although most people are satisfied with the transmission they get on most calls, there are still many calls which they do not like because the received sounds are not loud enough or because noise interferes with the conversations.
3. With the acoustic pressure ratios shown in Table I, and assuming that not too many calls have significantly poorer ratios and are not adversely affected by noise, there should be very few calls 5 or 10 years hence which people would consider poor.
4. People do not want the received sounds to be louder than they now get on connections between customers both of whom are within a mile or so of the same central office.

This is only part of the story, because beyond mere improvement in acoustic pressure ratio, people will want transmission to be still more pleasing and convenient in other respects.

Already many of them want to talk over the telephone without holding anything in their hands—and this hands-free method probably will spread, just as clutchless gear shifting is spreading in the automobile field. Perhaps, also, some years hence people will also want to see as well as to hear. And in addition to talking (and possibly seeing), there will undoubtedly be a whole host of new requirements for data transmission to computing machines, television of all kinds—fast, slow, color, and perhaps even 3-dimensional—and nobody knows what else.

But whatever comes, it is the job of the telephone industry to provide the means of transmitting the necessary information.

The Elements of the Problem

There are three main parts to the job of improving transmission and keeping costs low, i.e., instrumentalities, engineering, and maintenance. They are like the legs of a tripod, the soundness of the whole transmission structure depending upon how well each carries its load, and upon how well each is co-ordinated with the others.

In order to fashion an over-all transmission structure which will do the best practicable job at the lowest practicable

---

Paper 56-780, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 27, 1956; made available for printing April 27, 1956.

H. R. Huntley is with the American Telephone and Telegraph Company, New York, N.Y.

cost, a wide variety of instrumentalities must be available, and they must be capable of high-grade performance and be not too expensive.

But mere availability of high-quality and low-cost instrumentalities is not enough. They are only the bricks and mortar out of which a structure can be fashioned. The structure itself must be designed by engineers, and what they do with the instrumentalities they have and how well they prepare for instrumentalities to come are as important as the instrumentalities themselves. Regardless of the quality of the instrumentalities and the soundness of the engineering, the desired results will not be obtained unless the whole complex is so maintained that actual performance is very close to its design capabilities hour after hour, day after day, year after year.

These three parts of the problem will be discussed in turn.

LOOPS AND TRUNKS

Before discussing these parts of the job in detail it may be well to delineate the two main categories of the plant from the transmission standpoint because the problems and methods are quite different. These two main categories are loops and trunks as shown in Fig. 1.

All telephone connections involve two loops or circuits (usually pairs of wires) connecting customers' telephones to central offices (switching centers), with central offices connected together by trunks.

A large number of the calls are between customers served by the same central office or by different central offices in the same exchange area (usually a single community or closely related group of communities) which may be connected by direct trunks. On many calls there are two or more trunks in tandem between the central offices, and in a few of the more complex long-haul connections of the future there may be as many as ten trunks in tandem between the central offices. Historically, trunks have been divided into local and long distance, or intra and intercity, but from the transmission standpoint this division is rapidly losing significance.

In addition to the fact that there often are physical and electrical differences between loops and trunks, there is another difference which has important economic and service reactions. A loop is a circuit which is set aside for the exclusive use of a single customer or of a small group of customers who constitute the parties on a party line. On the other hand, a trunk is a circuit which is used by many different pairs of customers in succession.

Thus, in loops there is no (or very little) diversity. Whatever transmission condition exists on a particular customer's loop affects every call he makes or receives, but does not affect the great mass of customers. From the economic standpoint the absence of diversity means that each loop, and hence the whole loop plant, stands idle most of the time.

In reverse fashion there is a great deal of diversity in the use of trunks. The transmission conditions on a trunk or a trunk group are not exclusive to particular customers; customers may get widely different transmission on successive calls if the trunks which they use are different, and transmission conditions on each trunk can affect a large number of customers. From the economic standpoint, the diversity permits trunks to be used much more efficiently than loops, for the trunk plant stands idle a much smaller proportion of the time.

This difference in diversity results in a very large difference in the numbers of circuits in the two categories. Thus, there are about 25 million loops serving some 50 million telephones but less than two million trunks.

Instrumentalities

TELEPHONES AND LOOPS

The customer's telephone and loop form a unit, and the transmission which this combination provides is a function of both of its parts. The telephone is one of the most important in the whole gamut of transmission instrumentalities.

Telephones come in a wide variety of sizes, shapes, and colors. But the only reason why any customer has one at all is to talk over it. And improvements in telephones are such an important factor in improvements in transmission and in keeping costs low, that a brief historical résumé may be in order.

In the early years before electronic gear came into such wide use in the telephone plant, transmission losses were likely to be very high and the main transmission problem was to get enough sound out of the receiver that at least something could be heard. Consequently, the main emphasis in telephone set development was to obtain the highest practicable efficiency. The state of the art then was such that high efficiency could be obtained only by making full use of the mechanical resonance of receiver and transmitter diaphragms even though this meant some sacrifice in flatness of frequency response.

Table I. Progress in Making It Easier to Talk

Time	Approximate Ratio of Acoustic Pressure,* Talker's Lips to Listener's Ear	
	Typical Long Distance Call†	Typical Local Call†
25 years ago	45 to 1	30 to 1
Now	20 to 1	13 to 1
5 to 10 years hence	6 to 1	4 to 1

* Weighted for bandwidth, sidetone, etc.

† Individual calls may differ considerably from "typical" values. Present range is about ±3 to 1 from figures. Range was greater in past, will be less in future.

However, about 25 years ago the art had developed to the point where reasonable efficiency could be obtained without relying so much on mechanical resonance and a much flatter frequency response was practicable. Fig. 2 compares the sets in general use up to 1930 and the later design in general use about 1945. While in the 1945 set the efficiency at frequencies around 1,000 cycles was substantially lower than in the 1930 set, the increased efficiencies at other frequencies and the improvements in flatness of the whole band made the 1945 set talk better.

The 1945 set was flat enough that attention could again be directed primarily to increasing the efficiency. The success that has been attained in the past 10 years or so can be seen by comparing the 1945 and 1954 curves in Fig. 2.[4]

Improvements in bandwidth and efficiency are only part of the problem; there is much more to it, for example: each increase in either transmitters or receivers and associated circuitry increases the probability of crosstalk, and this has to be taken into account in engineering cables and open-wire transposition systems. Similarly, each increase in transmitter or receiver efficiency increases sidetone and the circuitry of the telephone has had to be improved to maintain a nice balance between having the telephone sound too "dead" and having the sidetone so loud that the talker lowers his voice too much.[5] Also, of course, each improvement in receiving efficiency makes any noise present on the line sound louder, and while the signal-to-noise ratio remains unchanged, the noise may be more annoying; improvements in transmitting efficiency tend to improve the signal-to-noise ratio. And since people do not want speech sounds to be much louder than they now get on connections between two customers, both of whom are close to the same central office, any increase in basic efficiency carries with it the need for incorporating means to reduce the efficiency when the telephones are on short loops.

In spite of the problems involved, however, large improvements have been made in telephones and associated plant and more will be made. Up until now efficiency and bandwidth improvements in telephones have come about mainly by improvements in tried and true components: the carbon-button transmitter and the moving-diaphragm receiver. While some further improvements may be possible along these same lines, it is not at all out of the question that the telephone of the future may be a quite different device. It seems quite likely that transistors will play a prominent part in future telephone design even though it is too early to tell precisely what part they will play.

Utilization of improvements in efficiency of telephones can obviously be directed toward either or both of two objectives: to improve transmission and to keep costs down.

For some years now, and for the foreseeable future, the economics of the overall situation are such that these improvements have been and will be devoted as far as practicable to control of costs. Primary responsibility for transmission improvements will be allocated to trunks, because the job can be done there more effectively and at less cost, as will be discussed later.

There are two ways in which set improvements can be used to help limit costs:

1. Finer loop gauges can be used. This effect can be seen from Table II where the percentage of 26-gauge loop cable has increased markedly in spite of the fact that the average length of loop has increased also.
2. Transmission limitations on number and locations of switching centers can be greatly reduced, thus permitting use of the most economical arrangements.

However, transmission improvements alone do not necessarily permit finer gauges or longer loops because it is also necessary to transmit signals from the customers' telephones to tell the switching gear what connections to set up and when to take them down, and from the switching gear to the telephones to tell the customers that somebody wants to talk with them. This is a component of signaling.

All through the history of loop design either signaling or talking has controlled the use of gauges for loops of different lengths. At the moment and in spite of the fact that signaling limitations have been greatly eased, loop plant is controlled by signaling, although not by a large factor.[6]

The next big step, therefore, must encompass both transmission and signaling improvements—and they are both being worked on intensively.

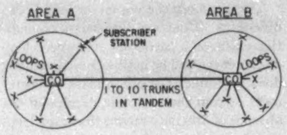


Fig. 1. Loops and trunks

Helpful as they may be, however, improved signaling and increased transmission efficiency will not completely overcome the basic economic problem of lack of diversity in the use of the loop plant. Something even more drastic is needed. There are two obvious ways to go about this problem, and starts have been made along both paths.

One method is carrier. While this method does not increase the diversity in the use of circuits, it does permit each pair of wires to carry several circuits and hence increases the diversity in the use of the physical plant. At present carrier is most attractive in the rural plant and several systems have been designed for such applications.[7–11]

The other method is line concentration, which extends the trunking principle of diversity to some point between the main switching installations and the individual customers. Line concentration has had only limited application so far, but it is under active study both here and abroad.

At the present time, the equipment involved is still quite expensive. But since the cost of equipment is almost entirely the cost of fabrication of components and assembling them into working circuits, the cost can—and will—be reduced as quantities increase, manufacturing techniques improve, and designs get better.

These things, together with improved telephones and signaling, provide much hope for the future.

TRUNKS

As indicated before, trunks must carry the major part of the burden of further improvements in transmission, as they have in the past. They are in a better position to carry this burden than are loops because there are so few of them relative to the number of loops, and they are usually longer.

Although long lengths increase technical problems and require almost fantastic precision in components, design, engineering, and maintenance, they do facilitate the use of electronic instrumentalities, upon which the promise of better transmission at low cost rests.

The two most widely used forms of electronic gear are voice-frequency repeaters (2-way audio amplifiers with associated circuitry), and carrier type (equipment which permits a pair of wires or coaxials to carry several or thousands of circuits). Radio is also a form of electronic gear, but in the telephone industry its principal application is as an alternative to wires and cables as a base facility for carrier. This subject will be discussed later.

Here it may be worth while to examine briefly an important difference between circuits which have electronic gear and those which do not. Once a circuit is equipped with electronic gear, the loss at which it is operated does not affect its cost. The loss at which it is operated depends solely on how the electronic gear is adjusted. The manner in which the electronic gear is adjusted depends entirely on technical factors and not at all on economic factors.

This is not true for a circuit without electronic gear. The only way to reduce loss on such a circuit is to use larger conductors or different physical arrangements, or heavier weights of loading, all of which increase cost. For such circuits, the longer they are the heavier the conductors must be so that the costs per circuit-mile increase with length.

While both carrier and repeaters have the advantage of divorcing cost and loss, carrier is particularly advantageous for the longer circuits, because it has high velocity of propagation and is basically 4-wire. In 4-wire operation, transmission in one direction travels on a different pair of wires or radio beam or in a different frequency region from that in the other direction, which greatly simplifies problems of echo, crosstalk, and singing. It would be difficult to overstate the transmission advantages of carrier for the longer circuits. For a given loss, a carrier circuit can be five or six times as long as the best voice-frequency cable circuit and over 20 times as long as the more common type of cable circuit.[13]

It is not too much to say that carrier is the most important single factor in providing good transmission on long-distance connections, and in permitting the telephone industry to proceed with its rather complex plans for industry-wide dial operation with full assurance that the customers can talk, and talk well, over the connections set up by the machines. Thus, for trunks of lengths where carrier is economical, the burden of providing low transmission losses is a very slight one. And, in fact, if it were economically practicable to use carrier in all trunks the

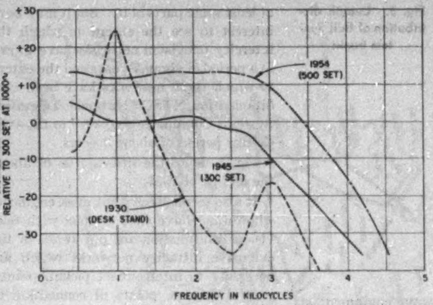


Fig. 2. Frequency response comparisons for telephone sets including two 3-mile 22-gauge loops

job of providing optimum transmission would be greatly eased.

However, carrier has the characteristic that while it reduces the cost per circuit of the base facilities (line plus carrier repeaters or radio beams), it does so only by using relatively complex electronic gear at the ends. The contribution to cost per circuit-mile of the terminal equipment is inversely proportional to length, whereas the contribution of the base facilities is independent of length. This situation leads to two rather distinct phases of carrier development.

One phase results in what might be called "long-haul" systems, which are primarily for use where the terminals are so far apart that their contribution to cost per circuit-mile is small as compared to the contribution of the base facilities. These systems are characterized by very efficient use of base facilities, made possible by relatively complex terminal equipment and elaborate regulating systems.[13–17] Probably the best examples of this line of development are the coaxial systems which are extensively used here and abroad. In North America two versions of the coaxial system are in use: the *L1* which provides 600 circuits on two coaxials (or two radio beams);[16] and the *L3* which provides 1,860 circuits on two coaxials (or radio beams).[17] Both are also suitable for television transmission.[16,17]

The other phase of development is directed toward the "short haul" systems. Here the emphasis is on producing the cheapest and most flexible terminal equipment practicable, so that the terminals can be close together without contributing extensively to cost per mile. Efficient use of base facilities is sacrificed to the extent necessary (but no more than this extent) to accomplish this aim.[14–20]

From the very start of carrier 35 or more years ago, there has been an urge to cheapen terminals, but only since World War II has it been possible to implement the short-haul concept effectively. Since World War II, developments have been very rapid and now there are many short-haul systems on the market. It would be difficult to single any one of them out for special mention, but it may not be too much to say that the well-known *N1* system had a great deal to do with setting off the chain reaction.[18]

The service and economic advantages of being able to use carrier in shorter lengths are pointed out in Fig. 3. As shown, over 90 per cent of the trunks in the Bell System are less than 25 miles long and over 80 per cent are less than 15 miles long.

Even with the modern short-haul carrier systems which are on the market, it is difficult to prove carrier in at distances below about 15 miles (at least in cable areas). This means that for over 80 per cent of the trunks some other method must be used. This is not necessarily a long-term condition because there is no reason to believe that the ingenuity of the industry in developing new types of carrier and in producing carrier cheaper has been exhausted. In fact, it is reasonable to expect that transistors and other new forms of components, together with increasing knowledge, will, if anything, accelerate the historic trend in favor of carrier vis-a-vis voice-frequency cables and wires.

While further reductions in costs of carrier, and probably further increases in costs of outside plant, will make carrier attractive economically for shorter and shorter lengths in the future, it is impracticable to wait until this happens. So, for a length of range under 15 miles, it is still necessary to use voice-frequency circuits, and the thing which has had a profound effect on this range is the modern version of the voice-frequency repeater, the negative-impedance design.[21–24]

Negative-impedance repeaters permit adding amplification in the voice-frequency range at such low cost that fine-gauge wire (usually 24 with a little 22 gauge) can be used for practically all trunks which are too short for carrier.

While reduction in conductor size has not offset the increase in cable costs over the past 25 years—or even the last 8 years—it has at least softened the blow so that trunk costs even in the shorter lengths have not gone up as rapidly as cable costs.

Although carrier would provide trunks of superior transmission even in these short-length ranges, the modern series and shunt negative-impedance repeaters supply an adequate transmission facility. Transmission losses can be reduced by using simple negative series impedance, but the use of series devices alone inherently produces "bumps" in the impedance-frequency characteristics of the circuits. The magnitude of these bumps is roughly proportional to the gain provided by the device. By combining with the series element a shunt negative-impedance element in a "T" configuration, this inherent limitation on the series element can be eliminated and the same kind of performance can be obtained as with the older type of repeater involving two amplifiers and two hybrid coils.

The increasing penetration by electronic gear into shorter and shorter length ranges is causing repeaters and carrier channels to be produced in quantities which would have been considered astronomical only a few years ago. For example, before World War II, repeaters were added in a few tens of thousands per year; now they are going in at a rate of over 100,000 per year. Before World War II, carrier channels went in at a few thousand per year; now they are going in at 30,000 to 50,000 per year. It is primarily upon this electronic gear that the promise of better, more convenient, and more pleasing transmission rests.

RADIO

The principal use of radio is as a base facility for carrier, but this does not mean that it is not a very significant development.

Table II. Exchange Cable Shipments

	Per Cent Conductor Miles		
Gauge	1930	1947	1955
19	17.6	11.4	7.1
22	30.3	25.9	25.0
24	46.5	33.9	33.5
26	5.6	28.8	34.4

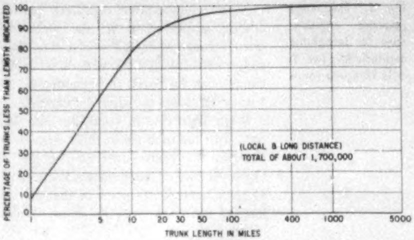


Fig. 3. Length distribution of Bell System trunks

Of the 33 million miles of telephone carrier circuits in the Bell System, almost 23 per cent are on radio relay. Of the 69,000 television channel miles, almost 70 per cent are furnished by radio relay.

The current heavy-duty radio-relay system is the *TD-2*,[25] which operates in the 4,000-mc (megacycle) band. In approximately 5 years *TD-2* has grown to the point where it has about 1.6 times as many route miles as coaxial cable. These routes carry almost 70 per cent as many telephone circuit-miles as coaxial cable and almost 2.3 times as many channel miles for television service. Usually, *L1* carrier channelizing is used on *TD-2* for telephone service.

Work is going ahead rapidly on an even heavier duty radio-relay system which will operate in the 6,000-mc band and with which *L3* carrier equipment will probably be used for telephone service. Like the *TD-2* it will be capable of high-grade color as well as monochrome televison transmission.

In the short-haul field, radio is used for only a comparatively small percentage of the total circuit mileage. This is partly because suitable systems have only recently become available and partly because as a base facility it must compete with existing base facilities (the very large existing networks of cable and open-wire lines) which still have some room for carrier. However, a number of short-haul systems of several different manufacturers are being used to provide regular telephone service, and a new system is being developed which will operate in the 11,000-mc band and which will be applicable for short-haul telephone and both monochrome and color television service. It seems very likely that the use of short-haul microwave systems will increase quite rapidly to care for growth in areas which are largely supplied by open-wire lines and to secure improved service reliability by diversifying routes and types of facilities.

The Bell System is also using a significant amount of microwave equipment of several manufacturers for television pickups, studio transmitter links and special services.

In addition to the use of radio as base facilities for carrier, the telephone industry provides an extensive mobile telephone service. The more than 15,000 cars, boats, and trains equipped with this type of service can make calls anywhere over the regular telephone network. The industry also installs and maintains both point-to-point and mobile radio facilities for customers' private use. Over 17,000 mobile sets are currently provided under this arrangement.

Other Forms of Transmission

Even though transmission on ordinary telephone calls (local and long-distance) is by far the largest quantitative factor in the transmission job, there are many other parts to the over-all problem. Some of these other parts involve speech transmission, but others involve significantly different forms of transmission. Some of the special setups for speech transmission pose very severe problems, but there is little basic difference from ordinary telephony and the instrumentalities involved are much the same, although they may be assembled and used in quite different ways.

Forms [illegible]smission other than speech include teletypewriter and much faster forms of pulse transmission (as for computers or other data-processing and utilization schemes), transmission of photographs, and television. It is obviously impracticable here to do more than mention one or two of these forms of transmission.[26-29]

TELEVISION

Most engineers are aware that the telephone industry furnishes an extensive intercity television transmission network and most of them are probably aware that color television transmissions occur over at least some parts of it. But it may be of interest to see the extent to which the intercity television networks have grown in a period of about 7 years and the extent to which these networks have been conditioned for NTSC (National Television Systems Committee) color[30,31] in the even shorter period of about 3 years.

Fig. 4 shows the situation as of about the middle of 1955.

It also seems likely that most engineers who do not have close contact with television transmission are not aware of the extensive intracity networks which are necessary to interconnect pickup points, control centers, points of connection to the intercity networks, television transmitters, etc. Fig. 5 shows the layout of special wire circuits used for this purpose in Manhattan. Each of the special circuits consists of a special polyethylene-insulated shielded pair together with suitably spaced amplifiers which are capable of precisely matching the loss and the propagation time of the special pair at all frequencies between about 30 and 4,000,000 cycles per second. In some of the complex switching arrangements used by the television industry, as many as four of these circuits may be connected in tandem between the pickup point and the connection to the intercity network and the compensation must be so precise that even under these conditions the loss must be correct to a few tenths of a decibel and the phase must be correct to a few degrees over the whole frequency range.

The switching arrangements involved pose quite a problem. Not only must they be capable of almost instantaneous major rearrangements of both the inter- and intracity parts of the networks on cue or signal (to care for different pickups, insertions of commercials, etc.), but they must do it without upsetting the precise attenuation and phase adjustments over the whole 4+-mc band.

Truly, the television transmission problem consists of much more than setting up a circuit to get a picture from one place to another.

At the other end of the television transmission scale is a rapidly growing requirement to transmit stationary (or very slowly moving and usually rather simple pictures for display on the face of a tube at a remote point. For example, the banking industry may want tellers at remote points to be able to compare signatures or check bank balances from central records; other industries may want to read meters or check numbers or other things at remote points.

Here, speed in transmission is much less important than in regular television,

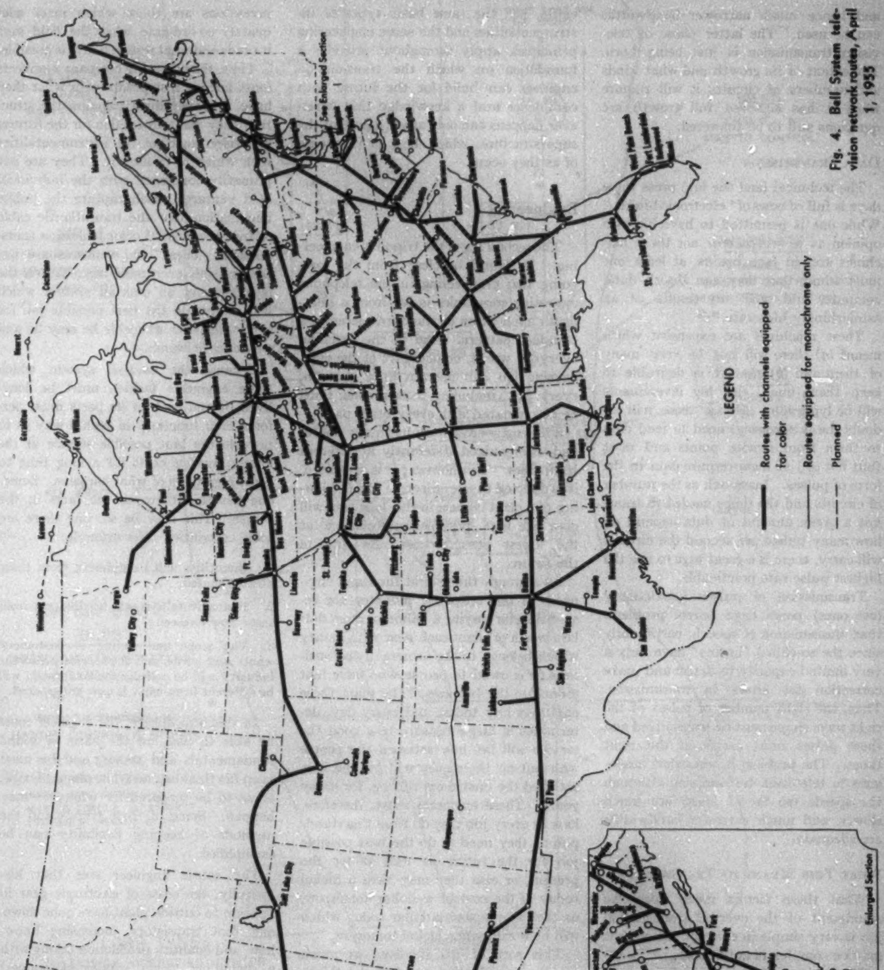

LEGEND
Routes with channels equipped for color
Routes equipped for monochrome only
Planned
See Enlarged Section
Enlarged Section
Seattle
Portland
San Francisco
Fresno
Bakersfield
Los Angeles
San Diego
Salt Lake City
Cheyenne
Denver
Colorado Springs
Albuquerque
El Paso
Phoenix
Tucson
Amarillo
Lubbock
Wichita Falls
San Angelo
Fort Worth
Dallas
Waco
Houston
San Antonio
Shreveport
Lake Charles
Beaumont
Oklahoma City
Tulsa
Wichita
Kansas City
Omaha
Lincoln
Sioux City
Sioux Falls
Fargo
Valley City
St. Paul
La Crosse
Green Bay
Milwaukee
Chicago
St. Louis
Springfield
Little Rock
Memphis
New Orleans
Birmingham
Atlanta
Jacksonville
Tampa
St. Petersburg
Miami
West Palm Beach
Fort Lauderdale
New York
Boston
Washington
Philadelphia

Fig. 4. Bell System television network routes, April 1, 1955

and hence much narrower bandwidths can be used. The latter phase of television transmission is just being born. The extent of its growth and what kinds and numbers of circuits it will require when it has achieved full growth are questions still to be answered.

DATA TRANSMISSION

The technical (and the lay) press these days is full of news of "electronic brains." While one is permitted to have his own opinion as to whether or not these machines are, in fact, brains, at least one must admit that they can absorb data, compute, and spill out results at an astonishingly high rate.[28,29]

These machines are expensive which means (*a*) there will not be very many of them and (*b*) that it is desirable to keep them busy, else big investments will be lying idle. Hence there will no doubt be a growing need to feed data to them from remote points and most (but not all) of them require data in the form of pulses. Inasmuch as the number of circuits and the times needed to transmit a given amount of data depend on how many pulses per second the circuits will carry, there is a great urge to use the highest pulse rate practicable.

Transmission of pulses (particularly fast ones) poses more severe problems than transmission of speech, particularly since the so-called "brains" have only a very limited capacity to detect and make correction for errors in transmission. Thus, the right number of pulses of the right wave shape must be transmitted and these pulses must arrive at the right times. The problem is somewhat analogous to television transmission although the speeds (so far at least) are much slower, and much narrower bandwidths are adequate.

WHAT THIS MEANS TO TRANSMISSION

What these factors mean from the standpoint of the over-all transmission job is very simple in concept although it may be complex in application.

The basic concept is that all forms of transmission involving a given bandwidth are improved or degraded by the same things. For example, quantitatively the things which can be done to a telephone circuit to improve transmission of data will also improve speech transmission; conditioning of a broad-band carrier system for color television will improve its performance if it is used for telephone.

It is true that quantitatively the requirements for different forms of transmission may be quite different. However, the fact that the same basic concepts, i.e., the same basic types of instrumentalities and the same engineering principles, apply throughout provides a foundation on which the transmission engineer can build for the future with confidence and a knowledge that whatever happens can only affect details of the superstructure, which he can take care of as they occur.

Engineering

The second leg of the tripod is engineering. It should be clear from the foregoing that engineering in the telephone operating companies is far from a humdrum application of standard pieces to a standard pattern. And in the practical everyday world, no one part of the engineering job can operate in a vacuum; every part, including transmission, must be co-ordinated with every other part.

This engineering work has many phases and is organized differently in different companies. But however it is organized, it is directed to one objective, the fashioning of a plant which, in the long run, will give the best possible performance at the lowest possible cost now and in the future.

As a group, the several thousand engineers in the telephone industry are responsible for buying a billion or more dollars worth of plant each year with money which belongs to the owners of the business or is owed to people who have lent money to the business. The plant these engineers buy today, and every day, determines in large measure how good the service will be, how satisfied the people who put up the money will be, and how satisfied the customers will be, for many years. These engineers must, therefore, look at every job they do from this standpoint; they need to do the best possible job for the future as well as for the present, or else they may save a nickel today at the cost of a dollar tomorrow, or they may spend a dollar today which will be worth only a nickel tomorrow.

This sort of job involves as much creative imagination and forward thinking as the development of the instrumentalities themselves. In fact, the creative imagination and forward thinking on the part of the operating company engineers must go beyond merely making the best possible use of what is at hand, but it must guide the development of the instrumentalities that are needed to do the future job.

It is still true that necessity is the mother of invention. And while it may not be true that research is the father, it is certainly true that the most valuable inventions are those which most adequately co-ordinate what the field says it needs and what research says is possible.

Thus, the operating company engineers must fashion a structure with what they have, and they must visualize the structure they want to fashion for the future, and they must get the instrumentalities with which to fashion it. They are not primarily concerned with the individual bold ventures which capture the public imagination like the transatlantic cable or transcontinental color television transmission. These bold ventures are important, but even more important is the fashioning of an over-all system which will always do the best possible job for every customer wherever he may be and whatever he wants.

Because the over-all system which these engineers fashion must be long-lived, it and each of its parts must perform their functions in such a way as to provide the best possible service at the lowest possible costs for a long time to come regardless of what happens. Someone has said there are no facts in the future. This may be so, but there are, some certainties. For example:

1. Quantities will be different from those now estimated.
2. Instrumentalities will be different from those now foreseen.
3. The scope and nature of customers' wants and needs, and the service which the industry will be called upon to furnish, will be different from what is now anticipated.

In this atmosphere, the engineer must be able to plan on the basis of sound fundamentals and trends, and he must keep his thinking, and his plant, flexible so as to be prepared for whatever may happen. Some of these trends and the methods of keeping flexibility can be exemplified.

The astute engineer sees that historically, the costs of electronic gear in relation to outside plant have gone down, and that transistors, increasing know-how, and quantity production (along with a dwindling world supply of copper) should accelerate this trend. He plans, therefore, on the basis of minimum physical plant and the maximum use of electronic gear.

He sees that improvements in switching gear, improved methods of handling traffic, etc., together with improvements in telephones and other transmission instrumentalities, have made radical changes in number, locations, and sizes of switching installations, and he knows that even more radical changes are in the offing. He again uses electronic gear to

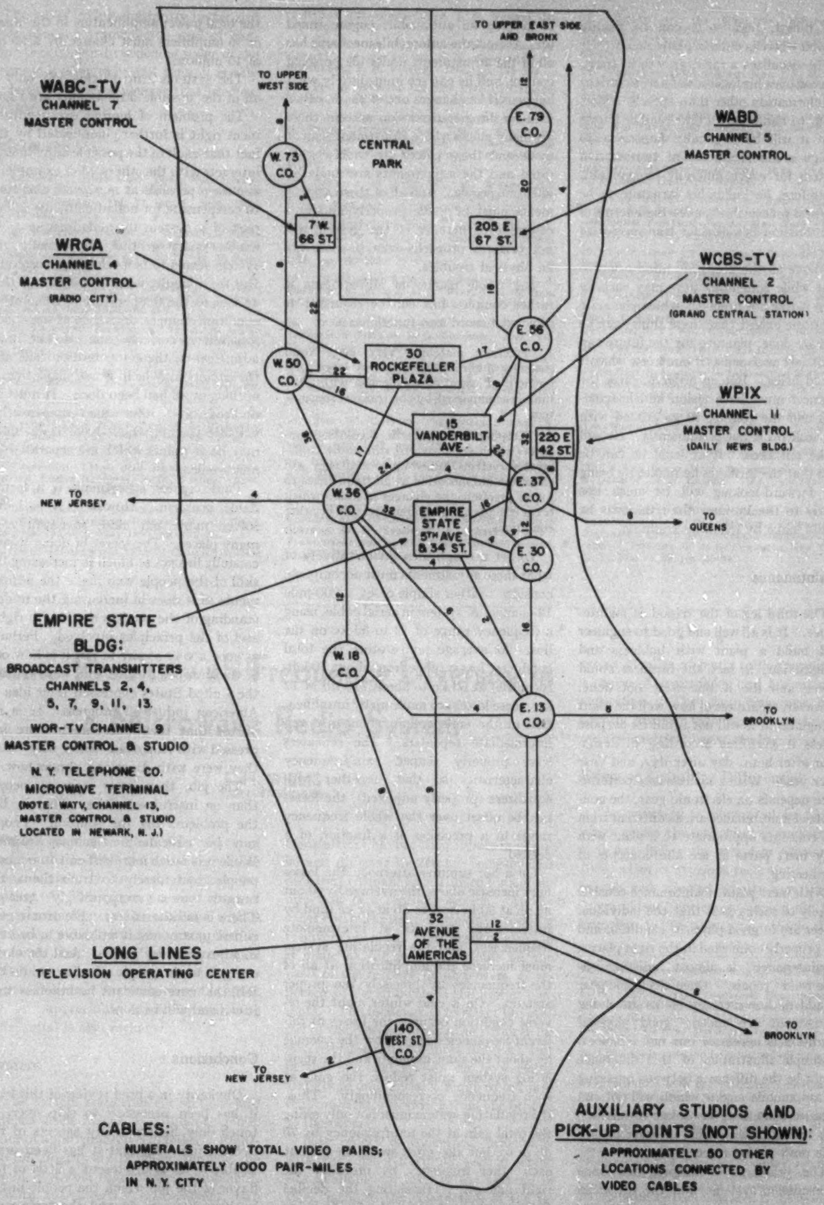

TO UPPER WEST SIDE
TO UPPER EAST SIDE AND BRONX
WABC-TV
CHANNEL 7
MASTER CONTROL
W. 73 C.O.
CENTRAL PARK
E. 79 C.O.
WABD
CHANNEL 5
MASTER CONTROL
7 W. 66 ST.
205 E 67 ST.
WRCA
CHANNEL 4
MASTER CONTROL
(RADIO CITY)
WCBS-TV
CHANNEL 2
MASTER CONTROL
(GRAND CENTRAL STATION)
E. 56 C.O.
W. 50 C.O.
30 ROCKEFELLER PLAZA
WPIX
CHANNEL 11
MASTER CONTROL
(DAILY NEWS BLDG.)
15 VANDERBILT AVE
220 E 42 ST.
TO NEW JERSEY
W. 36 C.O.
E. 37 C.O.
TO QUEENS
EMPIRE STATE 5TH AVE & 34 ST.
E. 30 C.O.
EMPIRE STATE BLDG:
BROADCAST TRANSMITTERS
CHANNELS 2, 4, 5, 7, 9, 11, 13.
WOR-TV CHANNEL 9
MASTER CONTROL & STUDIO
N. Y. TELEPHONE CO.
MICROWAVE TERMINAL
(NOTE: WATV, CHANNEL 13, MASTER CONTROL & STUDIO LOCATED IN NEWARK, N. J.)
W. 18 C.O.
E. 13 C.O.
TO BROOKLYN
32 AVENUE OF THE AMERICAS
LONG LINES
TELEVISION OPERATING CENTER
TO BROOKLYN
140 WEST ST. C.O.
TO NEW JERSEY
CABLES:
NUMERALS SHOW TOTAL VIDEO PAIRS;
APPROXIMATELY 1000 PAIR-MILES
IN N. Y. CITY
AUXILIARY STUDIOS AND PICK-UP POINTS (NOT SHOWN):
APPROXIMATELY 50 OTHER LOCATIONS CONNECTED BY VIDEO CABLES
8
4
8
22
12
20
4
18
17
22
16
8
36
17
24
4
32
22
32
8
4
16
12
4
4
4
12
6
2
16
4
8
8
6
12
2
4
2
8

Fig. 5. Video wire service: Principal cable routes and equipment locations in Manhattan, May 13, 1955

the fullest, because it can be readily moved whereas outside plant cannot.

He visualizes a rapid growth in transmission over his facilities of a wide variety of information other than speech—video, data, to name only a few—and he knows that it will be physically impossible to design a different form of transmission system for every different requirement. Therefore, he builds his structure to be the best practicable one for these forms of transmission as well as for transmission of speech.

Above all, he keeps his mind open, so that while something new may surprise him, it will not upset his whole plan.

To the extent that these things can be and are done, planning for the future can be almost as sound as if one knew what it would bring. But, in order to plan, the engineer must have vision and imagination and the courage to go around with his neck stuck out permanently. He will make mistakes, but at least he can be sure that the mistakes he makes by being too forward-looking will be much less costly to the business than the ones he would make by following tradition.

Maintenance

The third leg of the tripod is maintenance. It is all well and good to engineer and build a plant with boldness and imagination; in fact, the business would sicken and die if this were not done. However, regardless of how well the plant is engineered it will not fulfill its purpose unless it performs according to design hour after hour, day after day, and year after year. With a plant whose performance depends on electronic gear, the concepts of maintenance are as different from the concepts appropriate to a plant with only inert parts as are the concepts of engineering.

With inert plant, maintenance consists largely of seeing to it that the individual pieces are in good physical condition and are properly connected in the right places; "maintenance" is almost indistinguishable from "repair." Once electronic gear is added, however, problems involving adjustment are added; good physical condition is necessary but not sufficient. A simple illustration of this difference might be the difference between repairing an automobile engine which will not run because something is broken or missing, and tuning up an engine in order to provide peak performance.

The problems of keeping telephone transmission systems, and the telephone circuits they provide, tuned up is akin to but much more complex than the problem of keeping an automobile engine tuned up. At least the automobile mechanic has all of the adjustments under his personal control, and he can see immediately what happens if he changes one of them. But, in a far-flung transmission system there are many places where adjustments can be made and these places are widely separated and the adjustments are made by different people. And all of these adjustments must be made properly and in a co-ordinated manner or the system will not perform properly even if there are no physical troubles.

The whole matter of adjustments is rather complex but can be regarded as directed toward two functions:

1. Precisely matching very large total amounts of amplification to very large total amounts of loss at all frequencies within the range encompassed by the transmission system.

2. Keeping the automatic regulating systems (which are provided with most transmission systems) tuned up so that they will cause the amplification at all frequencies to follow precisely the changes in losses which occur due to changes in temperature or other conditions.

To get a rough idea quantitatively of what these adjustments must accomplish, consider a rather simple case: a 200-mile 12-channel *K* system in aerial cable, using a frequency range of 12 to 60 kc on the line. At average temperature the total conductor losses range from about 700 db (decibels) at 60 kc to about 310 db at 12 kc; these losses are made up by amplification at the ends and in about a dozen intermediate repeaters. The repeaters have properly shaped gain-frequency characteristics so that, together with equalizers (properly adjusted), the losses can be offset over the whole frequency range to a precision of a fraction of a decibel.

On a hot summer afternoon the losses may increase above the average by about 35 db at 60 kc and 15 db at 12 kc (and by intermediate amounts at intermediate frequencies), and the regulating system must increase the amplification at all of the frequencies by precisely the proper amount. On a cold winter night the reverse condition occurs; the losses at different frequencies drop below the average by about the same amounts and the regulating system must reduce the gain at each frequency correspondingly. Thus, the regulating system must not only swing the total gain at the top frequency by 70 db or so, but also must swing the gain at each other frequency by precisely the right amount. Translating the decibel figure to a power ratio, the 70-db swing between summer and winter means that the total power amplification in the dozen or so amplifiers must change by a factor of 10 million.

The systems can do this job only if all of the possible adjustments are right.

The problem of having every adjustment right is further complicated by the fact that each of the possible adjustments interacts with the others. For example, it is always possible at any particular time to compensate for maladjustments in one part of a system by maladjustments in another part so that superficially the system seems to be performing correctly. But the difficulty when this is done is that as soon as the temperature or some other condition changes, or as soon as somebody somewhere corrects one of the maladjustments, the compensation fails and the circuits may be in worse shape than if nothing at all had been done. It must be realized, also, that the compensating maladjustments which have been made may be at points which are separated by many miles.

Thus, correct adjustment is a formidable problem. However, it is being solved more and more successfully in many places. The secret of doing it successfully lies not so much in increasing the skill of the people who make the adjustments as it does in increasing the understanding of the need to do the job right, and of the principles involved. Perhaps as good a way as any to put it is how one of the European teams who were visiting the United States expressed their idea of American industrial progress. It is reported that they were much more impressed with American "know-why" than they were with American "know-how."

The job, then, rests more on people than on instrumentalities. Therein lies the problem, and the solution. People can be educated; machines cannot. While it is much more difficult to educate people than merely to train them, the rewards are correspondingly greater. There is no alternative. Electronic gear is here to stay and it will have to be kept in adjustment by people. And the closer one gets to the people who have to do the job, the more confident he becomes that it can and will be done.

Conclusions

Obviously in a brief review of this kind it has been necessary to skip over, or touch only lightly, many aspects of the over-all job. All that it has been practicable to do is to present a little of the flavor of the job which the people in the business have to do and the tools with which they have to work.

The job of giving constantly better and more pleasing transmission and of handling the constantly widening scope of the business is a big one; but then, it is only by meeting such challenges that progress is made.

References

1. Automatic Switching for Nation-wide Telephone Service, A. B. Clark, H. S. Osborne. *AIEE Transactions*, vol. 71, pt. I, Sept. 1952, pp. 245–48.

2. Transmission Design of Intertoll Telephone Trunks, H. R. Huntley. *Bell System Technical Journal*, New York, N. Y., vol. 32, Sept. 1953, pp. 1019–36.

3. A Revised Telephone Transmission Rating Plan, P. W. Blye, O. H. Coolidge, H. R. Huntley. *Ibid.*, vol. 34, May 1955, pp. 453–72.

4. An Improved Telephone Set, A. H. Inglis, W. L. Tuffnell. *Ibid.*, vol. 30, Apr. 1951, pp. 239–70.

5. An Improved Circuit for the Telephone Set, A. F. Bennett. *Ibid.*, vol. 32, May 1953, pp. 611–26.

6. Simplified Transmission Engineering in Exchange Cable Plant Design, L. B. Bogan, K. D. Young. *AIEE Transactions*, vol. 73, pt. I, Nov. 1954, pp. 498–502.

7. Telephone Lines for Rural Subscriber Service, Lester Hochgraf, R. G. Watling. *Ibid.*, vol. 74, pt. I, May 1955, pp. 171–76.

8. Objectives and General Description of the Type-P1 Carrier System, R. C. Boyd. *Ibid.*, vol. 75, pt. I, May 1956, pp. 188–91.

9. Application of the Type-P1 Carrier System to Rural Telephone Lines, J. D. Howard, Jr. *Ibid.*, pp. 205–14.

10. Circuit and Equipment Descriptions of Type-P1 Carrier System, E. K. Eberhart, F. J. Hallenbeck, E. H. Perkins. *Ibid.*, pp. 195–204.

11. Power Supplies for the Type-P1 Carrier System, D. H. Smith. *Ibid.*, pp. 191–95.

12. Some Recent Developments in Long Distance Cables in the United States of America, A. B. Clark. *Bell System Technical Journal*, New York, N. Y., vol. 9, July 1930, pp. 487–92.

13. Carrier is King, C. M. Mapes. *Bell Telephone Magazine*, New York, N. Y., vol. 28, Winter 1949–1950, pp. 191–203.

14. A Carrier Telephone System for Toll Cables, C. W. Green, E. I. Green. *Bell System Technical Journal*, New York, N. Y., vol. 17, Jan. 1938, pp. 80–105.

15. An Improved Cable Carrier System, H. S. Black, F. A. Brooks, A. J. Wier, I. G. Wilson. *AIEE Transactions*, vol. 66, 1947, pp. 741–46.

16. Progress in Coaxial Telephone and Television Systems, L. G. Abraham. *Ibid.*, vol. 67, 1948, pp. 1520–27.

17. The L3 Coaxial System—System Design, C. H. Elmendorf, R. D. Ehrbar, R. H. Klie, A. J. Grossman. *Bell System Technical Journal*, New York, N. Y., vol. 32, July 1953, pp. 781–832.

18. The N1 Carrier Telephone System, R. S. Caruthers, W. E. Kahl, L. Pedersen. *Ibid.*, vol. 30, Apr. 1951, pp. 1–32.

19. The Type O Carrier System, P. G. Edwards, L. R. Montfort. *Ibid.*, vol. 31, July 1952, pp. 688–723.

20. Combined Operation of 24-Channel Cable Carrier and 12-Channel Open Wire Carrier Systems, G. W. Searle, R. A. Schaefer. *AIEE Transactions*, vol. 74, pt. I, 1955 (Jan. 1956 section), pp. 770–77.

21. Theory of the Negative Impedance Repeater, J. L. Merrill. *Bell System Technical Journal*, New York, N. Y., vol. 30, Jan. 1951, pp. 88–109.

22. Theory of E-Type Repeaters, J. L. Merrill, Jr. *AIEE Transactions*, vol. 73, pt. I, Nov. 1954, pp. 443–47.

23. E-Type Telephone Repeaters, Description, Equipment and Testing, J. O. Smethurst. *Ibid.*, pp. 435–43.

24. Negative Impedance Telephone Repeaters—Application in the Bell System, A. F. Rose. *Ibid.*, pp. 430–35.

25. The TD-2 Microwave Radio Relay System, A. A. Roetken, K. D. Smith, R. W. Friis. *Bell System Technical Journal*, New York, N. Y. vol. 30, Oct. 1951, pp. 1041–77.

26. Telegraphy in the Bell System, J. A. Duncan, R. D. Parker, R. E. Pierce. *AIEE Transactions*, vol. 63, 1944, pp. 1032–44.

27. Carrier Telegraph System for Short Haul Application (43A1), J. L. Hysko, W. T. Rea, L. C. Roberts. *Bell System Technical Journal*, New York, N. Y., vol. 31, July 1952, pp. 660–87.

28. Transmission Line Characteristics and Effects on Pulse Transmission, P. Mertz. *Proceedings*, Symposium on Information Networks, Brooklyn Polytechnic Institute, Brooklyn, N. Y., Apr. 12–14, 1954.

29. Transmission of Digital Information Over Telephone Circuits, A. W. Horton, Jr., H. E. Vaughan. *Bell System Technical Journal*, New York, N. Y., vol. 34, May 1955, pp. 511–28.

30. A New Local Video Transmission System, S. Doba, Jr., A. R. Kolding. *Ibid.*, July 1955, pp. 677–712.

31. Intercity B-W and Color Television Transmission, J. M. Barstow. *Transactions*, Professional Group on Broadcast Transmission Systems—1, Institute of Radio Engineers, New York, N. Y., Mar. 1955, pp. 55–64.

Some Results with Frequency Diversity in a Microwave Radio System

F. H. WILLIS
NONMEMBER AIEE

ONE way to combat the effects of multiple-path fading of microwave radio signals is to use diversity reception.[1,2] This paper presents some of the transmission results obtained with a 2-frequency diversity system supplied by Motorola, Inc., to the Northwestern Bell Telephone Company and installed in North Dakota. An analysis of the effect of an appreciable diversity-switching differential is also included.

System

The system comprises six sections and extends from Minot to Williston, a distance of approximately 120 miles. Eighteen toll circuits are provided through the use of Motorola multiplexing equipment.

With the diversity arrangement, two 6-kilomegacycle frequency-modulation transmitters having their inputs in parallel and differing by 240 mc (megacycles) in output frequency are utilized for a given direction in a repeater section. At the other end of the section two corresponding receivers are used. The branch selection is made at the receiver outputs on the basis of the relative input powers, as reflected in the strength of the limiter currents. The tendency is to select the output of the receiver having the stronger input, over the range of most interest; but in the practical system an appreciable amplitude differential is required to cause the comparator to switch from one branch to the other.

This report is concerned principally with the Minot-Berthold section which was available for tests in the autumn of 1954, prior to the completion of the remaining sections. The test link extended across 23 miles of gently rolling terrain in a region of grain fields and low-growing vegetation. The ray clearance was slightly greater than first Fresnel zone on a geometrical basis; see Fig. 1.

Effect of Switching Differential

The situation is illustrated in Fig. 2 which shows the switching characteristic as it was during the period in November 1954, when continuous recordings of the branch and selected signals were made. In an "ideal" system, assuming the two branches to be alike, the switch from

---

Paper 56-1001, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956, and re-presented at the Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted July 25, 1956; made available for printing September 24, 1956.

F. H. Willis is with Bell Telephone Laboratories, Inc., New York, N. Y.

The author serves as spokesman for many people who contributed to the tests: in particular, Henry Magnuski and J. J. Sedik of Motorola, Inc., K. J. Frolund, R. L. Lindaas, and R. G. Pearson of the Northwestern Bell Telephone Company, and S. R. Collis of the American Telephone and Telegraph Company. The analysis of the recordings was performed by the author's colleagues, D. Gibble and B. S. Lamb. The approach used in the Appendix resulted from discussions with S. O. Rice.

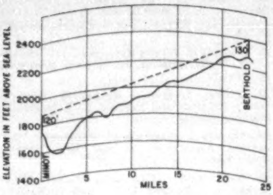
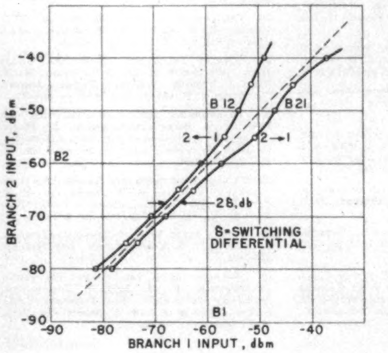


Fig. 1 (left). Profile of test repeater section

Fig. 2 (below). Diversity-switching characteristic in test section



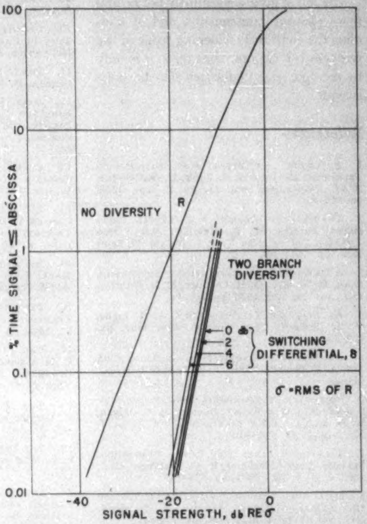


Fig. 3. Effect of switching differential on distribution of selected branch signal strength

branch 1 to branch 2 would occur as soon as the branch 2 receiver input B_2 exceeded that of branch 1, B_1, and vice versa. The transition would occur at the dashed line representing branch signal equality. In obtaining the measured characteristic, B_2 was held constant at a particular value and B_1 was varied, and the values B_{12} and B_{21} corresponding to switches 1→2 and 2→1 were determined. The ratio B_{21}/B_{12}, in db (decibels), is labeled 2δ; and δ for the symmetrical case is defined here as the switching differential. (The measured curves for different receiver combinations tend to flare out in the region of strong signals in a variety of shapes depending largely upon the saturation characteristics of the branch receivers.)

The differential is appreciable and may be a function of the strength of the branch signal being used. While some differential is desirable in that it may prevent needless switching on shallow fades or at other times when the branch fields are scintillating about a common value, it does exact a toll in the form of a decrease in diversity improvement from that obtained with zero differential. The effect is investigated in the Appendix. It is found that the degradation is by no means prohibitive in the case of moderate switching differentials. The analysis applies to the important region of deep fades subject to the condition that the switching differential is sensibly constant over this region and is not unreasonably large.

When the branch fading is of the Rayleigh type, the analysis leads to the approximate result in the form of an inequality applying when $as^2 << 1$ as in following equation:

$$s^4 < P(z<s) < (a^2+1/a^2)s^4/2 \qquad (1)$$

where $P(z < s)$ is the probability that the selected signal amplitude z will be weaker than a value s, where s is expressed as a ratio to the rms value of the branch amplitude distribution; and a is defined by $\delta = 20 \log a$, where δ is the switching differential, as given earlier.

In the inequality, equation 1, the value s^4 is just the distribution for the "ideal" system ($a=1$, $\delta=0$ db). The necessity of leaving the result in the form of an inequality at the present time (not a matter of practical concern, in view of the small degradations even at the upper value of the inequality, with representative switching differentials) is discussed in the Appendix. To extend the analysis beyond this would involve a statistical study of the crossings of the signal strengths through the switching region, making use, say, of simultaneous recordings of the branch signals. This would be an illuminating study, but it is not undertaken here.

The result, equation 1, is exemplified graphically in Fig. 3 where curves for the ideal case of zero-switching differential ($\delta=0$ db), and also for $\delta=2$, 4, and 6 db are plotted. The assumed individual branch distribution R (Rayleigh) is also included.

As an example of the application of the curves or formulas, refer again to Fig. 2, the switching characteristic applying to the test link. Here in the weak signal region of interest 2δ equals about 3 db, corresponding to a value of about 1.2 for a. Then, from the inequality, expression 1, the percentage increase in $P(s)$ over that with an ideal comparator would be not greater than six.

It is well to note that according to Kaylor the Rayleigh distribution may be a good approximation to the fading during severely disturbed periods, but there is sometimes a tendency for the rms value σ of the distribution to fall several decibels below the nonfading value of the signal.[2] This effect, when its magnitude is known or estimated, can be included in using the curves of Fig. 3.

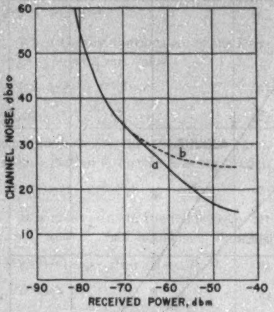


Fig. 4. Effect of fade in test section on channel noise

A—Test section only
B—Over-all noise in 6-section system

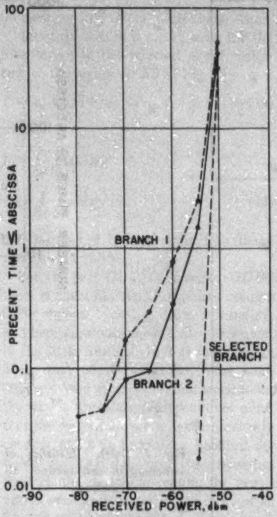


Fig. 5. Cumulative distributions for entire period November 6–28, 1954

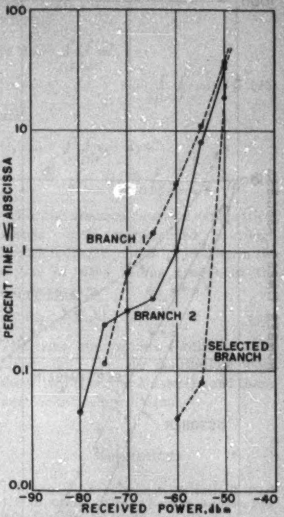


Fig. 6. Cumulative distributions, covering disturbed periods

Test Results

Noise Versus Received Power

The observed effect produced on channel noise by a fade in the test section was as shown in Fig. 4. The received power under nonfading conditions was about −48 dbm (decibels referred to one milliwatt). Branch noise, expressed here in decibels adjusted at zero transmission level, rises considerably on deep fades, and it is then that the diversity tends to be most effective. The trend for the over-all 6-section system as obtained in August 1955, some months after the installation had been completed is also indicated.

Diversity Improvement

By the use of three Esterline-Angus recorders and appropriate circuitry, the inputs of the two branch receivers (6,015 mc and 6,255 mc), and of the selected branch respectively, were recorded continuously over the period of November 6 to November 28, 1954. In obtaining the distributions, the recordings were divided into 5-db intervals.

The cumulative distributions derived from the charts covering the whole recording period are shown on Fig. 5. Those of Fig. 6 apply to disturbed periods defined somewhat arbitrarily for present purposes as starting one hour before the signal power fell below −55 dbm and ending one hour after the return to that value.

The disturbed periods amounted to about 20 per cent of the total time. The percentage would be greater in summer. The branch distributions might be expected to approach the Rayleigh during periods of severe multiple-path fading. The inclusion of some only slightly disturbed periods would move the distributions away from the Rayleigh in the direction observed. The large diversity improvement is apparent.

The cause of the apparently lesser fading on branch 2 than on branch 1 is not known. A longer period than the one analyzed would be required to determine the tails of the distribution curves with precision.

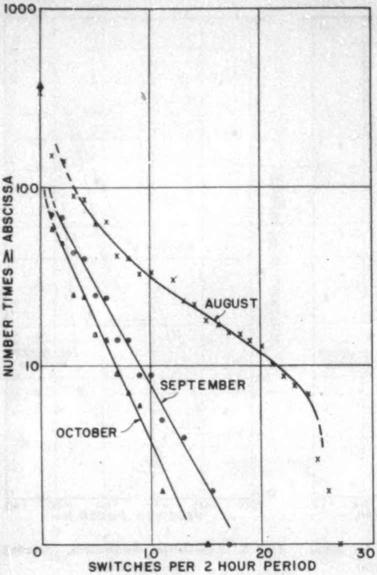
Frequency of Diversity Branch Switching

In connection with the transmission of certain kinds of information, there is a possibility that errors may result when a switch is made from one branch to another, and therefore a knowledge of the frequency of occurrence of the switching is desirable. For this reason, the switching occurrences were recorded in several sections of the system during parts of the month of July through October 1955.

In analyzing the data, the time was divided into 2-hour periods, starting at midnight each day, and the number of switches in each period was counted. In addition, the number of switches for each 5-minute interval was counted during periods of "large switching activity." These periods were arbitrarily defined as any hour containing more than eight switches. Cumulative distributions of the observed switching rates for the Minot-Berthold link are shown in Fig. 7 and Fig. 8. Whether the rates are excessive or not for a particular service depends on the degradation which a switch imposes on that service and upon the requirements.

Conclusions

The remarks which follow apply to periods of active fading of the kind encountered in the trial: 1. The effect of a practical switching differential is appreciable but by no means prohibitive. 2. With a 240-mc frequency separation the branch fadings tend to be uncorrelated. 3. Large noise improvements are realized by the use of the diversity arrangement. The use of diversity in a representative light-route system comprising a few repeater sections and meeting typical toll noise objectives increases the effective margin against selective fading by some 15 db. However, diversity reception is ineffective against field strength depressions caused by a blocking or deflection of the rays from the region of the receiving antenna during certain tropospheric conditions which occur in some climates.



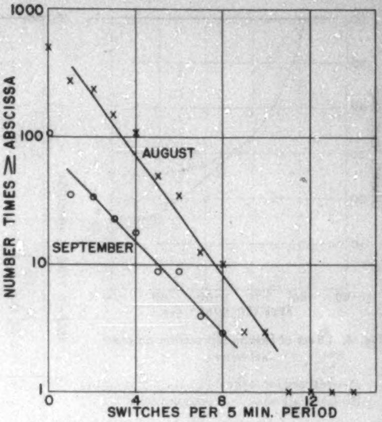


Fig. 7 (left). Statistics of switching in test section, all periods

Fig. 8 (above). Statistics of switching, covering active periods

Appendix

The results for an ideal 2-branch diversity system which selects the stronger signal and suppresses the weaker are well known, at least for the case of uncorrelated fading.[1] In the following the treatment is extended to the situation in which an appreciable amplitude differential δ db, say, is required to switch from one branch to another.

Ideal Diversity Reception

It is convenient to start with the notation and procedure of reference 1. Let x, y, z represent the respective amplitudes of the branch signals X, Y, and the selected signal Z at any time, at the receiver inputs. Let $\varphi(x,y)\,dx\,dy$ be the probability that the X and Y signals will fall simultaneously in the intervals x, $x+dx$ and y, $y+dy$ respectively, and let $p(x)\,dx$ and $p(y)\,dy$ be the individual probabilities.

If the comparator chooses the stronger signal, with $\delta=0$, the probability integral for the selected signal will be

$$P(z<s)=\int_0^s\int_0^s \varphi(x,y)\,dx\,dy \tag{2}$$

In the important case in which the branch fluctuations are statistically independent, $\varphi(x,y)=p(x)\ p(y)$ and, by equation 2

$$P(z<s)=P(x<s)P(y<s) \tag{3}$$

If, in addition, the branch distributions are equal, the following obtains

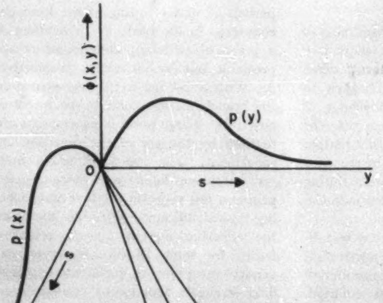


Fig. 9 (left). Switching characteristic assumed in analysis

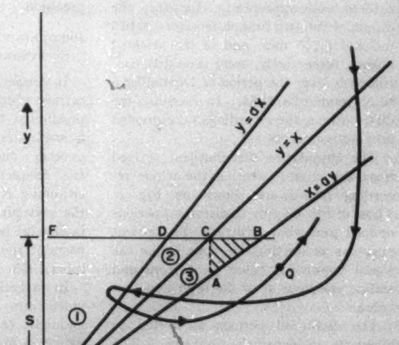


Fig. 10 (right). Representative signal strength point Q crossing the switching region

$$P(z<s)=[P(s)]^2 \qquad (4)$$

When the distributions are of the Rayleigh type with rms value σ

$$p(x)=(2x/\sigma^2)\epsilon^{-x^2/\sigma^2} \qquad (5)$$

$$P(x<s)=1-\epsilon^{-s^2/\sigma^2} \qquad (6)$$

and similarly for the Y branch. Then, from equation 4, for the selected branch

$$P(z<s)=(1-\epsilon^{-s^2/\sigma^2})^2 \qquad (7)$$

It is convenient in the following to let $\sigma=1$, and in that normalized notation

$$P(x<s)=1-(1-s^2+s^4/2!-\ldots) \qquad (8)$$

$$\approx s^2 \text{ when } s^2\ll 1 \qquad (9)$$

Similarly

$$P(z<s)\approx s^4 \text{ when } s^2\ll 1 \qquad (10)$$

Inclusion of Appreciable Switching Differential

The treatment applies to the weak signal region ($s\ll 1$). In that region δ in decibels is approximately uniform with signal strength, and for present purposes the situation can be represented by Figs. 9 and 10. The selecting switch will be either on X or Y and for the symmetrical case assumed will tend to be on each for half of the total (long) time.

When the switch is on Y, write for the distribution of the selected signal, referring to the region $OFBO$ of Fig. 10

$$P(z<s)=2(P_1+p_{231}\times P_{23}) \qquad (11)$$

where

$$P_1=\iint\limits_{OFDO}\varphi dx\, dy \qquad (12)$$

$$P_{23}=\iint\limits_{ODBO}\varphi dx\, dy \qquad (13)$$

and p_{231} is the factor required to make equation 11 valid.

For the part $ODCAO$ of region $ODBO$ there is equal likelihood of points coming from regions 1 and 4, from symmetry, and here p_{231} would equal 1/2. However, in the little triangle ACB it seems more likely that points will have come from region 4 than from 1. This depends on the way in which the representative point Q of the system moves, whether in slowly curving paths or in kinks. Lacking detailed information on the crossings through the switching region, the result nevertheless can be left in the form of an inequality satisfactory for present purposes

$$I_1<P(z<s)<I_2 \qquad (14)$$

where

$$(1/2)I_1=\iint\limits_{OFDO}\varphi dx\, dy+(1/2)\iint\limits_{ODCAO}\varphi dx\, dy \qquad (15)$$

and

$$(1/2)I_2=\iint\limits_{OFDO}\varphi dx\, dy+(1/2)\iint\limits_{ODCBAO}\varphi dx\, dy \qquad (16)$$

For the assumed case of a symmetrical system with uncorrelated Rayleigh branch fading, equation 14 reduces in the region of deep fading ($s\ll 1$) to the approximation

$$s^4<P(z<s)<(a^2+1/a^2)s^4/2 \qquad (17)$$

The same approximate result, by symmetry, holds for the distribution when the switch is on the X branch, and therefore for the selected signal.

References

1. Diversity Reception, Z. Jelonek, E. Fitch, J. H. H. Chalk. *Wireless Engineer*, London, England, Feb. 1947, pp. 54–62.

2. A Statistical Study of Selective Fading of Super High Frequency Radio Signals, R. L. Kaylor. *Bell System Technical Journal*, New York, N. Y., Sept. 1953, pp. 1187–1202.

Noise Considerations on Toll Telephone Microwave Radio Systems

T. A. COMBELLICK
ASSOCIATE MEMBER AIEE

M. E. FERGUSON
ASSOCIATE MEMBER AIEE

THE NATION-WIDE toll-dialing network is creating an increasing demand for more toll circuits. The demand may be satisfied by the addition of any of several types of facilities if they conform to the over-all network requirements.

This paper is presented to show how microwave radio equipment can be analyzed to determine if, when integrated into a system, it will meet the over-all noise requirements of the nation-wide network. Radio system noise is divided into its basic components, and minimum values are suggested for each component. Finally, the manner in which the components combine to produce the measured noise in a channel is shown by an example.

In instances where the breach in language between the radio engineer and the telephone engineer is considered a deterrent to common understanding, attempts have been made to bridge the difference.

Noise Objectives

Most radio manufacturers specify noise performance in signal-to-noise ratio terms that are easily related to test-tone-to-noise measurements. However, in multichannel telephone communications systems, the ratio of noise to speech in individual channels is somewhat tenuous.

In speech transmission, it is customary to measure noise performance by comparing the interfering effect of channel noise to the interfering effect of an arbitrary reference noise. This is done with a special meter consisting of a volume indicator and weighting network that takes into account the frequency and transient responses of the transmission line, the telephone handset, and the human ear. The unit of interfering effect is the dba (decibels adjusted).[1]

In this paper, all noise values are in dba as measured with the Western Electric *2B* noise-measuring set or equivalent, using F1A weighting at the 0-db transmission level point. Transmission level is measured in decibels referred to a common point in the system or circuit. The reference point is usually the transmitting toll switchboard. The transmission level at the reference point is 0 db.[2]

Current telephone practice is to engineer toll facilities so that they will meet the requirements for the nation-wide toll-dialing network. A commonly stated objective for an end-to-end connection through the network is 38 dba. For example, this objective might be for a connection between New York and San

---

Paper 57-59, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 29, 1956; made available for printing November 20, 1956.

T. A. Combellick and M. E. Ferguson are with the Lenkurt Electric Company, San Carlos, Calif.

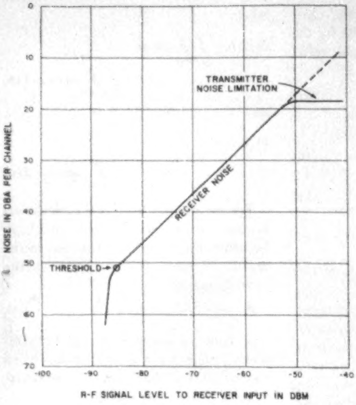


Fig. 1. Typical plot of system noise as a function of receiver input signal

Francisco, built up of five separate links by automatic switching equipment along the way. While seven links on one connection are possible under the toll-switching plan, the great majority of calls will be completed via five links or less. A link may be composed of wire, cable, or radio facilities, or combinations of these.

For the total noise of five links to be less than 38 dba, each link must have a noise objective of 31 dba (38–10 log 5). If the link is entirely one facility, all 31 dba may be assigned to it. If it is heterogeneously composed, the noise on each facility must be kept low enough so that the resultant composite of the link meets the objective.

Objectives of noise and crosstalk are normally stated on a statistical basis. Thus, the values given are for an average condition. Noise performance will degrade to be poorer than average some of the time and better than average some of the time. A reasonable limit for the worst degradation is that the noise on each link will not go below 51 dba for more than 0.1% (per cent) of the time.

Noise Sources

There are three principal sources of noise in a radio link. The first is the channelizing or multiplex equipment, the second is the transmission medium (the coaxial transmission line, waveguide, antenna systems, and the transmission path through the atmosphere), and the third is the radio equipment. This paper is primarily concerned with radio equipment noise. Channelizing equipment noise is somewhat familiar to most telephone engineers and can be easily measured by connecting the two channelizing terminals on a back-to-back basis. The channelizing equipment should be designed so that it does not contribute appreciably to the over-all noise in a radio link. A reasonable value to expect from this source is 20 to 23 dba.

Noise in the transmission medium is generated by echos or multiple paths of transmission so that the radio equipment receives two or more signals which may vary in phase. Noise originating in the transmission path when minimized by proper system engineering should not be a cause of poor performance for short- or medium-length systems. In frequency-modulated radio links of over 500 megacycles, the problem of man-made and atmospheric disturbances which produce noise in the radio receiver may be ignored.[3,4]

Radio Equipment Noise

The noise generated by the radio equipment may be divided into three categories that are relatively easy to measure or calculate. These divisions are transmitter noise, receiver noise, and intermodulation noise. It is important to note that interrelationships exist among them.

TRANSMITTER

Transmitter noise is primarily a function of the modulator or klystron tube used in a system and is caused by random current fluctuations in this stage. It is measured by connecting a radio transmitter and receiver back to back through a lossy line, and measuring the noise in a 3.5-kc slot in the baseband output under conditions of no load on the radio equipment. The value of transmitter noise is indicated when the measured noise does not decrease with further increases in received r-f (radio-frequency) signal strength. This is shown graphically in Fig. 1 as the top "leveling-off" point on the noise-versus-received-signal-strength curve.

Transmitter noise is not caused by intermodulation since it is present even when no signal is applied to the transmitter. The effect of transmitter noise can be minimized by increasing the modulation level and thereby producing more signal in the receiver so that the ratio of transmitter noise to transmitted intelligence will be reduced. However, increasing the modulation levels will also increase the intermodulation noise, so that a compromise between the noise sources must be achieved.

RECEIVER

Noise generated in the radio receiver is sometimes referred to as thermal noise. A theoretically perfect receiver would have an input noise power given by the familiar equation

$$\text{noise power} = KTB$$

where K = Bolzman's constant; T = temperature, Kelvin; B = bandwidth.

Since K is a fixed quantity and T is relatively constant, usually assumed fixed at 300 K (degrees Kelvin), the equation shows that the noise generated in a theoretically perfect receiver is determined by the receiver bandwidth term B. This is bandwidth ahead of the detector and includes the over-all r-f and i-f (intermediate-frequency) bandwidths.

All microwave receivers add some noise to that theoretically present at the receiver input. The additional noise is generated primarily in the r-f amplifier (if one is used), in the r-f mixer crystal, and in the first i-f stage of amplification. The ratio of this noise to theoretical minimum input noise determines the noise figure F_r of the receiver. It is usually expressed in decibels. Thus a noise figure of 9 db means that the receiver is 8 times noisier than the theoretical minimum possible with that bandwidth.

The theoretical minimum noise and the noise figure of the receiver determine the threshold point of the receiver. The threshold of an f-m receiver is the point at which the received signal power is just adequate to overcome the effect of noise-power peaks and thereby capture the receiver.

The received signal power required for this capture of the f-m receiver is generally accepted to be 10 db over the total receiver noise power ($KTB+F_r$), where KTB is the theoretical noise power for a given bandwidth, and F_r is the receiver noise figure expressed in decibels. For example, the noise power in a 5-megacycle band at 300 K is −107 dbm (decibels below one milliwatt). If a receiver has a noise figure of 12 db, the total receiver noise power is then −95 dbm and the receiver threshold is −85 dbm. In this particular receiver, when the signal level falls below −85 dbm the channel noise becomes very great, and when the received signal strength increases above −85 dbm the channel noise decreases linearly decibel for decibel until some limiting condition is reached. The limiting condition is usually caused by either intermodulation or transmitter noise.

It is always desirable to maintain as large a margin as possible between the received signal and the threshold point. During fades the channel noise increases rapidly as threshold is approached. Below threshold, the circuit becomes useless.

The lower the threshold noise, the greater the percentage of time the channel noise will be acceptable. The probability of a deep fade is smaller than that of a shallow one. The threshold point should be at least as low as the r-f signal level which is exceeded 99.9% of the time. The per-channel noise at threshold will then be equal to the noise which cannot be exceeded for more than 0.1% of the time. On many radio paths the limiting factor in determining the permissible locations of terminal or repeater equipment is the amount of average signal and the fade margin required.

Reduction in threshold level can be accomplished by making the bandwidth as narrow as possible and by reducing the receiver noise figure. The receiver noise figure is minimized by careful design of the receiver front end. The receiver bandwidth should be merely wide enough to pass the intelligence with a minimum of intermodulation distortion.

In selecting microwave radio paths the value of fades anticipated and the percentage of the time in which they occur can be determined on a theoretical basis.[5,6] Using such calculations, the transmission engineer tries to establish paths over which predicted depth and frequency of fades will fall within established system objectives. A system engineered to meet the usual link objective would have a predicted r-f signal input above the value corresponding to 51 dba for all but 0.1% of the time, or about 8 hours per year.

Systems designed in this manner usually have very low receiver noise under average propagation conditions. Under these average conditions, transmitter and intermodulation noises are the primary ones, while the receiver noise is important only during fades.

Above threshold, receiver noise varies inversely with r-f signal level on a decibel-for-decibel basis. In a system with both transmitter and receiver noise sources, the minimum noise is controlled by the transmitter noise. The system noise as a function of receiver r-f signal level can be plotted conveniently to form a curve such as shown in Fig. 1.

INTERMODULATION

The third source of noise in the radio equipment is intermodulation resulting from amplitude or phase distortion in the transmitter and receiver. Intermodulation distortion occurs when the system is loaded, and increases when the load increases. Since it is uneconomical to design for complete freedom from intermodulation, good engineering practice allows the intermodulation distortion to reach some small but finite value during a small percentage of the time.

On a system handling a load consisting of many speech channels, studies have shown that the complex multichannel speech signal is random in nature and can be simulated quite accurately by random noise. Because of the random nature of the load, the intermodulation products on a large channel system, 48 channels or more, are also random and result in a random-type noise on all of the channels.

Methods of measurement of intermodulation noise involve a determination of the level of the multichannel speech load to be expected. General practice has been to determine this load in terms of the peak loads which are exceeded during 1.0% of the busy hour.[7,8] The random noise equivalent of the speech load is then used to load all of the baseband except for a slot the width of a channel, usually 3.5 kc. The amount of noise appearing in this 3.5-kc slot at the receiver output is then a measure of intermodulation noise which will be exceeded for only 1.0% of the busy hour.

Increasing the modulation level to offset transmitter and receiver noise is desirable, but results in increasing system load and intermodulation distortion noise. A suitable compromise is to make the intermodulation noise, which is exceeded 1.0% of the busy hour, equal to the transmitter noise. If this balance of noise contributions is made, the per-channel noise under normal propagation conditions, i.e., when receiver noise is negligible, will increase less than 1.0 db for 90% of the busy hour and less than 3 db for 99% of it because of intermodulation.

Actual laboratory tests made to determine the proper modulation levels involve setting up a system with a large r-f signal input to the receiver to ensure that receiver noise is negligibly small. Without loading, the noise power in the 3.5-kc slot is measured. The noise generator which loads the baseband, except for the 3.5-kc slot, is then turned on and the noise loading power is adjusted until the noise in the 3.5-kc slot has increased 3 db over the unloaded value. The proper transmission level to the transmitter baseband input can then be readily determined in terms of the number of channels and the power of the noise signal which produces the 3-db increase.

Balancing Noise Contributions

It is evident that considerable interdependence exists among the various noise sources and the factors of bandwidth and modulation level. For optimum system design these factors must be balanced so that the noise requirements under average conditions and under 0.1%-of-the-time conditions are satisfied.

Since the receiver noise figure is always minimized, the principal factor determining the threshold point of the receiver is the bandwidth. Lowering the threshold point by reducing the bandwidth will cause increased intermodulation noise. Likewise, if the bandwidth is kept fixed, increasing the modulating level will decrease the effect of transmitter and receiver noises for any r-f input above threshold, and will cause increased intermodulation noise.

The proper balance between bandwidth, modulation level, and intermodulation noise is obtained when: 1. The threshold is kept as low as possible. 2. The per-channel noises at threshold equal the 0.1%-of-the-time requirement, 51 dba. 3. The 1.0%-of-the-busy-hour intermodulation noise equals the transmitter noise.

The previously stated objectives for noise are applicable on a link basis. When establishing the objectives for individual sections, it is necessary to know the number of sections to be expected in the link and also the way in which noise products combine. For a transcontinental link such as the Bell System's *TD-2* microwave system, the section noise objectives must necessarily be more stringent than for a short link.

The way in which different noise sources combine to form a resultant prod-

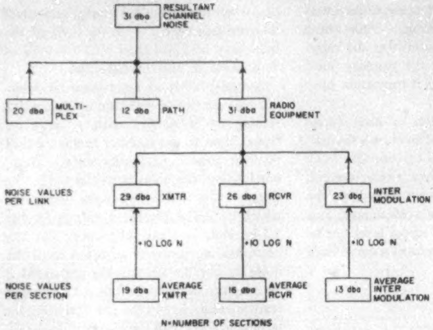

31 dba
RESULTANT CHANNEL NOISE
20 dba
MULTI-PLEX
12 dba
PATH
31 dba
RADIO EQUIPMENT
NOISE VALUES PER LINK
29 dba
XMTR
26 dba
RCVR
23 dba
INTER MODULATION
+10 LOG N
+10 LOG N
+10 LOG N
NOISE VALUES PER SECTION
19 dba
AVERAGE XMTR
16 dba
AVERAGE RCVR
13 dba
AVERAGE INTER MODULATION
N=NUMBER OF SECTIONS

Fig. 2. Channel noise under average conditions for 10-section medium-haul radio link

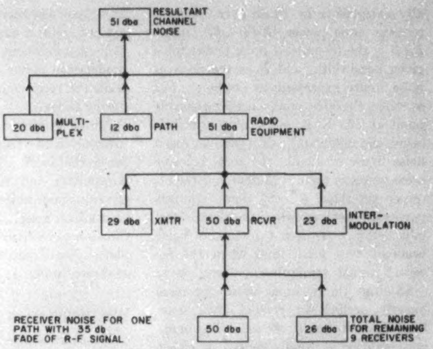

51 dba
RESULTANT CHANNEL NOISE
20 dba
MULTI-PLEX
12 dba
PATH
51 dba
RADIO EQUIPMENT
29 dba
XMTR
50 dba
RCVR
23 dba
INTER-MODULATION
RECEIVER NOISE FOR ONE PATH WITH 35 db FADE OF R-F SIGNAL
50 dba
26 dba
TOTAL NOISE FOR REMAINING 9 RECEIVERS

Fig. 3. Channel noise under 0.1%-of-the-time conditions for 10-section medium-haul radio link

uct is dependent on the nature of the noise. Pure random noise such as receiver and transmitter noise will always add to produce random noise, and the combinations add as power functions. This is also true of intermodulation when the second-order products are predominant. In multiplex systems where the baseband is contained in a single octave, second-order products cannot normally produce interference, but higher order products can. In these systems third-order intermodulation distortion predominates. The third-order per-section intermodulation products add in phase, and must be added on a voltage rather than on a power basis.

Fig. 2 illustrates the way in which the noise sources in a 10-section radio link will combine under average conditions for a radio system where second-order intermodulation products predominate. The factor of 10 times the logarithm of the number of sections is added to the average section noise for each noise source to give the resultant link noise from each source. The transmitter, receiver, and intermodulation noise factors are then added to give the noise attributable to the radio equipment. Finally, the radio equipment noise, channelizing equipment noise, and path noise are combined to give the resultant channel noise per link.

Fig. 3 illustrates the same system under an assumed 0.1%-of-the-time condition of a 35-db fade in one path. The resultant noise under these conditions increased by 30 db.

With reference to Fig. 2, it may be noted that under the 1.0%-of-the-busy-hour conditions the value for intermodulation will change from 23 dba to about 29 dba. This causes a 2-db degradation of the resultant channel noise for the link.

By using the same type of charts, knowing the number of sections, and starting with the link objectives, it is relatively easy to establish the values of noise required for each block. Radio and multiplex equipment manufacturers can usually supply specifications on noise contributions by their equipment. Using these data, the requirements for average receiver noise can be calculated and integrated into the radio-path requirements.

References

1. Noise Measurements on Telephone Circuits, **H. C. Franke.** *Bell Telephone System Monograph 2410*, Bell Telephone Laboratories, Inc., New York, N. Y.

2. Levels and Powers in a Carrier System. *The Lenkurt Demodulator*, Lenkurt Electric Company, San Carlos, Calif., vol. 1, no. 10, Dec. 1952.

3. Interference in Multi-Channel Circuits, **L. Lewin.** *Electrical Communication*, New York, N. Y., June 1951.

4. Intermodulation Noise in VHF Multi-channel Telephone Systems, **B. L. Slow.** *A.T.E. Journal*, Automatic Telephone & Electric Company, Liverpool, England, vol. II, no. 3, July 1955.

5. Radio Propagation at Frequencies Above 30 Megacycles, **Kenneth Bullington.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., Oct. 1947.

6. UHF Radio Relay System Engineering, **J. J. Egli.** *Ibid.*, Jan. 1953.

7. Load Rating Theory for Multi-Channel Amplifiers, **B. P. Holbrook, J. T. Dickson.** *Bell System Technical Journal*, New York, N. Y., vol. XVIII, Oct. 1939, pp. 624–44.

8. A Speech Volume Survey on Telephone Message Circuits, **V. Subrizi.** *Bell Laboratories Record*, New York, N. Y., Aug. 1953.

The Logical Design of a Digital Computer for a Large-Scale Real-Time Application

M. M. ASTRAHAN
ASSOCIATE MEMBER AIEE

B. HOUSMAN
NONMEMBER AIEE

J. F. JACOBS
NONMEMBER AIEE

R. P. MAYER
NONMEMBER AIEE

W. H. THOMAS
NONMEMBER AIEE

THE Lincoln Laboratory and International Business Machines Corporation (IBM) have, during the past 3 years, worked out the design for a new digital computer which is the central component of a large-scale real-time system. In this system, data from many sources are fed automatically into the computer where they are processed under programmed control. A complete compilation of the real-time situation is made by the computer and presented to operators by means of a special display system. The computer automatically generates control commands for the external environment in response to corrections and command information fed into it by the operators.

The purpose of this paper is to describe the performance criteria of the computer in general terms, and to present some of the outstanding features of the design. These features are necessary in the computer because of its particular real-time application which requires a greater emphasis to be placed on reliability, speed, capacity, and flexibility than is usual in scientific or commercial applications.

Disregarding its special features for the moment, the computer is a large-scale general-purpose single-address parallel digital computer with a 32-bit word length. A high-speed magnetic-core memory is provided which contains 270,336 bits of storage arranged in two banks of 33 planes, each plane consisting of a 64 by 64 core matrix. The 33d plane is used for parity check bits. In addition to its buffer drums for communicating with the external environment, an auxiliary drum storage system is provided. This system contains 3,244,032 bits of storage divided among eight cylinders, each consisting of six fields of 33-bit words. Each field has 2,048 words distributed around the circumference of a cylinder. Five magnetic tape units of the IBM *728* design and a large cathode-ray tube display system are also provided. In order to obtain a reliability capable of providing continuous 24-hour operation, the whole machine, with the exception of some of the input-output equipment, is duplicated. This guards against overall system catastrophies caused by sudden machine failures and allows planned partial shutdowns for maintenance purposes.

One major consideration in the design was the availability of components, particularly in the case of the high-speed memory. The original thinking on this computer resulted in a need for a memory faster and larger than in any existing computers. This pointed toward the magnetic-core memory development which had been under way for some time at Massachusetts Institute of Technology (MIT). The logical design of the computer was centered around the engineering judgment that the minimum memory cycle time would be in the order of 6 μsec (microseconds), and that the largest matrix of cores should be 64 by 64.

The availability of components also affected the choice of the high-capacity storage medium for which magnetic drums were chosen. The drum which seemed the most attractive and which was chosen was that used on the IBM *650*. The vacuum tubes and circuitry in the computer were developed from designs previously made by the digital computer laboratory at MIT and the Electronic Data Processing Machine development laboratories at IBM. The redesign of these basic circuits was dictated by the increased reliability requirements. In many cases this necessitated the use of more vacuum tubes and other components than are required in scientific or commercial applications.

The features of the computer which are the primary subject of this paper are in the logical design areas. The majority of these features were dictated by the need to obtain the most efficient use of the drum capacity and core-memory speed with a minimum number of circuits. As a result, special attention was given to overlapping operations within the machine, and to the balance between the available components and the logical arrangements which would maximize their usefulness.

The first of these features discussed is the arithmetic element. Many of the data which are processed by the machine are in Cartesian co-ordinates. Many of the computations which are involved in the application perform the same operation on x as on y. A saving in time is effected by the use of a dual arithmetic element which treats these quantities separately and simultaneously. The second feature is a high-speed multiplication technique which dictated the design of the adder in the arithmetic element. It was found that all of the proposed arithmetic operations, with the exception of multiplication and division, would be performed during the cycle time of high-speed memory. A special adder was developed so that the multiplication process could be made more compatible with the memory cycle.

The third feature is an indexing system which automatically takes care of the large class of bookkeeping operations, i.e., address modification, without taking extra operation time. The fourth feature is an input-output control system which was designed to allow the computer to utilize the time which might otherwise be wasted during periods when data are being transferred between it and the input-output system. This is accomplished by allowing the computer to continue doing work while waiting for the input-output equipment to present or receive its information. The fifth and final feature is a buffer drum system which matches the computer speed to the incoming data rate.

Arithmetic Element

In the design of a digital computer, much consideration must be given to the characteristic known as "word length." For many real-time applications, much of the input data are generated by analog devices such as thermometers, and much of the output data are sent to analog devices such as valves or indicators. A precision of 10^3 is often sufficient for such input and output data, with a precision of about 10^5 used to reduce the effects of roundoff and truncation errors during

---

Paper 57-200, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21-25, 1957. Manuscript submitted May 2, 1956; made available for printing December 17, 1956.

M. M. ASTRAHAN, B. HOUSMAN, and W. H. THOMAS are with International Business Machines Corporation, Poughkeepsie, N. Y., and J. F. JACOBS and R. P. MAYER are with Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass.

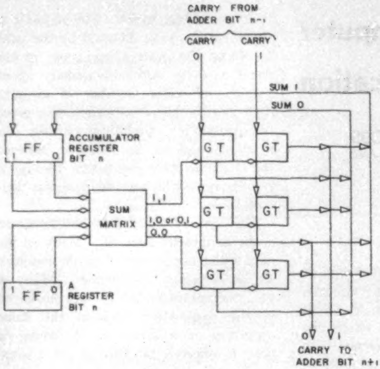


Fig. 1. Adder circuit

intermediate calculations. Thus a word length of from 10 to 16 bits is implied.

In a stored-program computer, consideration must also be given to the number of bits required to specify an instruction. A basic operation code of 64 instructions requires six bits. For memory sizes ranging from 2,048 to 16,384 registers, the address part of an instruction requires from 11 to 14 bits. Special control bits, such as those used for automatic indexing, add four or five more bits to the instruction word. Thus, the word length for a full instruction (including the basic operation code, control bits, and addresses) might be between 21 and 25 bits.

The conflict between the 16-bit word length for data and the 25-bit word length for instructions was resolved by the choice of a 32-bit word length. This allows two 16-bit numbers to be placed in any register, and allows additional control bits to be included in instruction words. These additional control bits are used to increase the potential speed of the machine. Since the memory must be capable of delivering the full register simultaneously for instructions, it can also deliver both numbers simultaneously. A separate arithmetic unit was provided for each of these numbers so that both could be manipulated simultaneously, thus doubling the speed of the computation. This arithmetic unit was termed the "dual arithmetic" unit.

Special operations and control bits have been added to the basic operation code to increase the utility of the dual arithmetic unit. A conditional transfer of control, or branch, can be made to take place if a specified arithmetic unit is negative, or only if both are negative. It is possible to store the full 32-bit word, or either half, independently. (The address part of an instruction is placed in the right-half word; thus a "right store" instruction can also be used for modifying addresses). Shift and cycle instructions are provided for shifting either arithmetic unit itself, or both together, or for cycling data between the right and left arithmetic units. A "twin" feature, available on some instructions, allows the left-half word from memory to be sent to both arithmetic units simultaneously. Thus, a vector in the dual arithmetic unit may be multiplied by a scalar from a left-half word of memory by a single "twin and multiply" instruction.

Each instruction which can cause an overflow contains a pair of control bits which allow suppression of an overflow alarm from either or both arithmetic units. Of course it is not always possible to make efficient use of the dual arithmetic element. Single 16-bit words can be handled just as easily by using the aforementioned controls, with no loss in speed except for an occasional cycle to exchange left for right.

Regardless of the word length used, there is usually a sizable body of data consisting of items only a few bits long. Storage space can be saved by packing a number of these items into a register of storage. Packing and unpacking such items can be very time consuming unless special instructions are given. In this computer, an "extract" instruction allows a given item to be obtained from a memory register without obtaining other items from the same register. A "deposit" instruction allows the item to be replaced in storage without disturbing other items in the same register.

High-Speed Multiply

The memory unit of a general-purpose computer must be used one or more times for every instruction. In a real-time computer the memory should be made as fast as possible and should never have to wait for other processes. In this computer, with the design goal of 6 μsec per memory cycle, it would have been desirable to design a 6-μsec multiplier because of the frequent use of the "multiply" instruction. On the other hand, a basic flip-flop speed of 0.5 μsec and a word length of 16 bits allowed a reasonably simple 8-μsec multiplier, using 0.5 μsec for each addition and shift.

Each step of a multiplication should allow for the possibility of a carry propagated a full register length, or 16 bits. Although this carry condition cannot occur mathematically, it can be used to simplify the design procedure because it is a reasonable approximation to mathematically possible carries. If the basic gate-tube propagation time is 0.04 μsec, then a 16-bit carry will require 0.64 μsec. To reach the goal of 0.5 μsec per multiplication step, the carry is accomplished concurrently with the 0.5-μsec flip-flop resolution time. This is done by making sure that all pulses which arrive at a given flip-flop are spaced 0.5 μsec apart, even though they are carry pulses which have arrived as much as 0.64 μsec after the initiation of the carry.

As shown in Fig. 1, each bit position makes use of a diode matrix to form the sum of two bits; one from the augend in the accumulator, and one from the addend in the *A* register. The matrix controls two sets of gate tubes; one set to be sensed by an incoming 0-carry pulse, and the other by a 1-carry pulse. When a carry pulse arrives, only one gate tube passes it and indicates the sum of the two bits and the carry. This pulse is then sent out as either an 0-carry or a 1-carry pulse to the next highest bit, and as either a 0-sum or 1-sum pulse to the accumulator flip-flop, depending on the value of the sum. Thus, the carry takes only the gate-tube propagation time to pass each bit. Carry pulses supplied to the right (the least significant) end at 2 megacycles will cause every bit to receive carry pulses (of either 0 or 1 value) at a 2-megacycle rate, although the pulses at each bit will be displaced in time, depending on the bit's distance from the right end of the register. This principle forms the basis of the so-called "asynchronous adder."

For multiplication, each "add" step must be accompanied by a shift right of

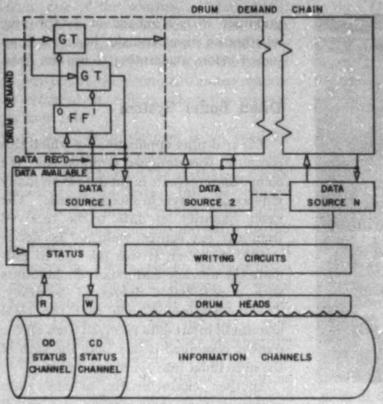


Fig. 2 (left). OD side, drum system

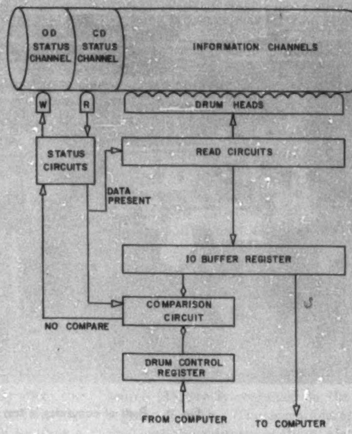


Fig. 3 (right). CD side, drum system

one place, and some steps (for a zero multiplier bit) will require a shift right without any addition. The shift on add is easily accomplished by sending the sum pulse from the adder to the accumulator flip-flop to the right instead of to the flip-flop which generated the sum. The shift right without addition must propagate down the accumulator at the same rate as the carry in order to avoid conflict with a previously initiated carry. This so-called "ripple shift" is accomplished simply by letting the shift pulse from one bit initiate the shift command for the next bit.

Special controls could have been included to allow the sum to be displaced to the left instead of to the right during a divide process, and to be undisplaced during a plain addition. However, the expense of such controls was not justified. A corrective shift left for add instructions (including a second corrective shift left after an end-around carry) may be accomplished without delaying the memory cycle. Unlike multiply, which is used extensively in most programs, the extra time required for "divide" does not appreciably increase the average instruction execution time because of the infrequent use of the divide instruction in most programs.

Indexing System

Most digital computer programs involve sequences of instructions which are repeated many times on different data. It wastes memory space to include separately stored instructions to process each piece of data since the instructions will vary only in the address to which they refer. On the other hand, time is required to modify instruction addresses for each execution. An indexing feature has been included in this machine to minimize this time.

The indexing feature consists of a set of four index registers, an index adder, and associated control circuitry. Each instruction which refers to a memory address may have control bits inserted into it to specify an index register. The contents of the specified index register are added to the address part of the instruction prior to its execution; however, the instruction as stored in memory remains unchanged. This indexing addition does not require any extra time since it is executed while waiting for the memory cycle to be completed. Thus, if an add instruction with an address part containing 1,000 is executed, and if its control bits specify an index register which contains 24, the instruction will be executed as if it had an address part of 1,024 but will remain stored in memory as add 1,000.

In order to utilize an index register in a cyclic program consisting of a "loop" of instructions, it must be possible to set up the index register to an initial value. Two instructions, "reset index register" and "reset index register from right accumulator," have been provided in this computer. They load the specified index register with the address part of the reset-index-register instruction or with the contents of the right accumulator.

There are three other functions that the indexing system must provide: 1. the modification of the index register each time the loop is executed, 2. the testing to determine if the loop has been executed the desired number of times, and 3. the branching of control back to the beginning of the loop unless it has been executed the desired number of times. All of these functions are handled by one powerful instruction, "branch and index." Control bits included in the branch-and-index instruction specify an index register and a decrement. Each time it is executed, the specified index register is first inspected to determine if it contains a negative number. If the number is positive, the contents of the register are reduced by the specified decrement and by the instruction branches to the address specified in its address part, usually to the beginning of the loop. Thus, execution of the loop will continue until the index register contents have been reduced to a negative number, at which time the branch is not executed and the program continues in sequence.

Besides the four index registers, the right accumulator may also be used as an index register. This feature was included to facilitate table look-up programs. Assume that a program has been executed so that the table argument has been computed in the right accumulator. If the next instruction executed is a "clear and add" instruction which specifies the right accumulator as an index register and has the first address of the table as its address, the first address and the argument will be added before the instruction is executed. Thus, with only one instruction, the desired table contents are obtained.

MEMORY UNIT LOCATION
MAINTENANCE CONSOLE
DRUM UNIT
TAPE UNIT
MEMORY UNIT

Fig. 4. Parts of computer in test cell

When an indexed loop is used in searching for a desired value in a table, it is often necessary to determine the contents of the index register when the desired value is found. An instruction called "add index" has been provided to permit this important operation.

Input-Output Control

The system application requires that large quantities of data be entered into the core memory, processed, and delivered out again. The data sources and destinations have varying and unpredictable word rates and access times. Since computing time is at a premium, the input-output control design goal was that an arbitrarily sized block of words be transferred with a minimum of time used for the transfer. This minimum consists of the time required to execute the program steps which set up the transfer, and the one memory cycle that is required to transfer each word into or out of core memory.

Reading into core memory will be assumed in the following: Writing out of memory is analogous. The design goal required the use of an independently operating input-output control which could remember the input-output unit selected, the number of words to be transferred, and the location in core memory for the block of words. A "break" system was needed which could interrupt the program operation for one memory cycle whenever the input-output unit had provided a word. The instructions in the computer are designed in such a way that a break memory cycle can be initiated at the end of any memory cycle with no effect on the instructions other than to delay their execution.

The system chosen operates as follows: Special instructions in the computer connect the proper input-output unit to the information transfer paths, load counters which keep a record of the location in memory in which the data are to be stored and keep a count of the words transferred, and start the flow of data. The program operation then continues normally, except for interruptions of one memory cycle per word caused by the break system. Each time a word is transferred, the counters are stepped accordingly in preparation for the next word.

The operation is terminated when the counters signal that the requested number of words has been received. It may also be terminated earlier by a disconnect signal from the input-output unit. This occurs when the program has requested more words than the input-output unit has to send and when the input-output unit has run out of words. Facilities have been included to allow the computer to examine the counters in order to determine how many of the words requested were transferred.

Interlocks hold up the program if an input-output operation is called for before the preceding one is finished. Therefore, to use the system efficiently, the programmer must provide enough work for the computer in order to consume all the time taken by the input-output operation before another input-output operation is started. To help accomplish this function, a "conditional branch" instruction is provided which can detect whether an input-output operation is still in process. This instruction also provides the programmer with a means of determining whether an input transfer has been completed before attempting to use the data.

Drum Buffer System

The real-time application for which this computer was designed required that the computer receive its input data from many independent asynchronous sources. The exact quantity of data to be received from any one source could not be determined; however, it was possible to estimate the average and maximum amounts with a reasonable degree of accuracy. The application is such that the total amount of input data received from all the sources combined is less than the sum of the individual maximum amounts.

Another characteristic of the data sources is that they operate at a much lower speed than that of the computer. It is not operationally feasible to interrupt the computer operations to accept each piece of data as it arrives. A buffering mechanism is necessary to gather the data at the slow incoming rate and then pass on large blocks of the data to the computer at the computer's speed. Magnetic drums were chosen for this buffer.

The desired characteristics of the drum buffer system follow: The input data should be written on the drum as soon as possible after they are received and definitely before another piece of data is received from the same source. The writing of input data on the drum should not interfere with the reading of data by the computer. The computer should be able to read selectively from the drum; i.e., to read data from only one source at a time. In order to minimize the number of drum storage registers required, the buffering system must be able to combine the data received from many sources.

To provide the "no interference" characteristic, a dual-access drum is used. This drum has two sets of drum heads. One set, called the outside of drum system (OD), is used to communicate only with the outside world. The other set, called the computer side of drum system (CD), communicates only with the computer.

To provide the "minimum storage" and "write as soon as possible" characteristics, the "random storage" drum was developed. Each drum register has a status bit associated with it. This bit is used to indicate the "full" or "empty" status of the register. As a register passes under the OD drum heads, the status of the register is sensed. If the register is empty, a "drum demand" pulse is generated and sent to interrogate the data-input equipment. If there are data from one of the sources, a "data available" pulse is sent back to the drum status circuits, and the data are sent to the drum-write register. As the information is written on the drum, the status marker is changed to indicate that the register is now full.

A single drum channel is not used for the status indication. There is not enough time to read the status bit, decide whether it indicates full or empty, generate the drum-demand pulse, interrogate the data sources, and write the full indication with the same drum head. If a second head were used on the same drum channel, for physical reasons, it would have to be located many registers away, thus necessitating the use of a shift register or delay-line device between the two heads. Therefore, two drum channels are used for the status indication, and the indicator bit is shifted back and forth between them. The two channels are called the OD and CD status channel since they are read by the OD and CD side of the drum respectively.

Consider first the OD side of the drum system as shown in Fig. 2. When a piece of data is received by the input equipment that is associated with a data source, a "data received" pulse is generated and is used to set the corresponding flip-flop in the drum-demand chain to the "one" condition.

As each drum register approaches the drum-write heads, the status bit on the OD status channel associated with the register is read. The status read head is located slightly before the write heads. If the status bit is a "one", indicating that the register is full, the only action that takes place is the writing of a "one" in the CD status channel by the status write head. If the status bit is a "zero," indicating an empty register, a drum-demand pulse is generated and sent to the drum-demand chain. This pulse interrogates each of the flip-flops in order until it finds one which has been set to the "one" state or until it reaches the end of the chain. If the pulse encounters a flip-flop in the "one" state, it becomes a data-available pulse and 1. resets the flip-flop, 2. causes the data from the source associated with the flip-flop to be transferred to the drum-writing circuits and 3. causes the status circuit to write a "one" in the CD status channel. If there are no data available, the "zero" indication is passed on to the CD status channel. Each piece of data written on the drum has some identity bits associated with it to identify the source of data.

There is a possibility that as the drum fills up, some of the data received will not be stored on the drum before the next piece of data is received from the same source. This is particularly true of the sources at the tail end of the demand chain. The probability of storage on the drum depends upon the rate of input data and upon the number of empty registers on the drum. Most of the data rates are such that if the drum is kept at least 50 per cent empty by the computer, the probability of storage is better than 0.99. The data received from this type of source are such that the occasional loss of a piece of data does not seriously affect the system. Fortunately, the rate from the critical sources is slow enough to allow a complete search of the drum for an empty register before a second message can be received from the same source.

Now consider the CD side of the drum as shown in Fig. 3. The status heads and circuits on the CD side are almost identical to those on the OD side. During periods when the computer is not reading the drum, the status information is automatically and continuously transferred from the CD status channel to the OD status channel.

The computer may read the drum in either of two modes, status or identity. When reading in the status mode, the contents of every full register passing under the heads is transferred to the core memory through the input-output buffer register and a "zero" is written in its associated status bit in the OD status channel, indicating that the register is now empty. When reading in the identity mode, the computer first places the identify code desired into the drum control register. Only those words with matching identity are transferred to the cores memory.

The transfer of words continues until disconnect pulse is received from the computer, indicating that it has received the number of words it has asked for, or until the drum itself generates a disconnect pulse. With an unknown quantity of data received from each source, the computer may ask for more data than has been received. In fact, in order to ensure that the computer receives all the data that have collected on a drum field between readings, the normal procedure is to ask for more than expected. A disconnect pulse is automatically generated at the end of a complete drum revolution, and the number of words read is determined from the word-counter contents.

Some of the data sources have messages which contain more information than will fit into one drum register. To handle such information, another feature has been added to the drums associated with these sources: the drum has been divided into multiple word slots of adjacent registers. For this application, the only meaningful status bit is the one associated with the first register of a slot, and the source identity is contained in the first register of a slot. The operations associated with those drums are almost identical to those of the single register drums.

In addition to the input buffer drums just described, there are output buffer drums which handle outgoing data. They operate in a similar, though inverse, manner to the input drums.

Conclusions

Fig. 4 shows a part of the computer in the test cell. Because of its size and layout, it was not possible to obtain a picture of the whole computer. The unit in the foreground is the operator's maintenance console. It contains switches for manual data or instruction entry and manual control, neon indicator lights for the major flip-flops and registers in the computer, visual and audible indicators for computer generated alarms, marginal checking controls and indicators, and power system and air-conditioning indicators. The inset shows one of the memory units.

The computer, including the directly connected input-output equipment, contains approximately 12,500 tubes. It has an execution time of 12 μsec for arithmetic instructions, excluding multiply and divide, which require 15.5 and 53 μsec respectively. The prototype model has been in satisfactory operation for more than 1 year.

Steady-State Operation of the Push-Pull Saturable-Core Transformer with Resistive Load

R. E. ANDERSON
ASSOCIATE MEMBER AIEE

Synopsis: The widespread use of the saturable-core transformer as a push-pull magnetic amplifier in servomechanism systems is well known. This amplifier is analyzed by the methods of Storm[1] to obtain the basic performance equations and to gain a broad insight into the circuit operation. The analysis is predicated on the free flow of even harmonics, that is, low control-circuit impedance, and deals with the resistive load case. Under these conditions two modes of operation are found; further, the influence of core loss and primary resistance on gain becomes apparent. From this analysis ideal graphs are prepared which show the nature of circuit operation for changes in signal, line voltage, frequency, and loading. The resulting expressions for gain, together with the boundaries of circuit operation, form a useful background in component design, quality control, and system analysis.

ONE OF the magnetic amplifiers which has received widespread usage during the past decade is the device known as the saturable-core transformer. This push-pull amplifier has been employed in numerous applications including the control of 2-phase induction motors in servomechanisms shown in Figs. 1 and 2 as modulators. The purpose of this work is to analyze the steady-state operation of the saturable-core transformer to explain the amplifier operation, and to obtain performance equations. Not only should such information have application value, but such equations would give broad insight into the design problem.

Thus consideration will be given to the analysis of the steady-state operation of the saturable-core modulator with resistive load. This device, as indicated in Fig. 1, consists of two identical transformers having series-connected primaries and series-opposition-connected secondaries, together with isolated control windings which produce saturation by direct current. In this paper the scope will be limited to transformers with ideal core materials and with low control-circuit impedance, that is, free flow of even harmonic currents.

Problem Solution

The problem with which this paper is concerned is to obtain expressions for the gains and outputs as a function of circuit parameters for the various conditions of applied power and of applied signals in the circuit of Fig. 2. Accordingly, for purposes of analysis the signal is considered as being applied to saturable-reactor unit A while unit B operates as a transformer loading. For purposes of analysis the polarities and current directions given in Fig. 2 will be used.

NOMENCLATURE

See Figs. 2, 16, and 18 for circuit locations.

e = instantaneous voltage, volts
E = average voltage, volts
E_X = X winding voltage on 2-unit basis, volts
f = power supply frequency, cycles per second
i = instantaneous current, amperes
I = average current, amperes
k_f = form factor, no units
K = gain, no units
m = slope of curve
n = matching index, no units
N = number of turns per core leg
P = power
R = resistance, ohms
t = time, seconds
T = time constant, seconds
V = voltages, volts
$\underset{\sim}{X}$ = per-unit excitation, no units
α = firing angle of saturable-reactor unit, electrical angle
β = saturating angle of transformer unit, electrical angle
Δ = incremental value
π = a number = 3.1416
φ = magnetic flux, maxwells
ψ = flux linkage, weber-turns
ω = angular frequency of power supply, radians per second

Subscripts

A = unit A
B = unit B
C = control circuit
G = gate winding on an individual leg basis
H = secondary winding (output circuit)
i = initial value
L = left leg
M = magnetizing branch
M = maximum or peak value
O = output circuit
R = right leg
S = saturation
TH = Thévenin equivalent
X = primary or input winding
Δ = swing, maximum minus minimum

Numbers

0 = zero output

1 = initial cyclic position
1 = mode 1 operation
2 = mode 2 operation

Those detail symbols which are not shown in circuit location are explained as follows:

E_{BGX} = B-unit induced voltage in X gate winding per leg, equation 2
E_{SX} = primary saturation voltage, 2-unit basis, equation 47
E_{XM} = peak value of primary voltage, equation 20
N_C = control winding turns per leg
N_H = secondary winding turns per leg
N_X = primary winding turns per leg
R_{HA} = H-winding resistance in unit A, equation 19
R_{HB} = H-winding resistance in unit B, equation 19
R_{TH} = Thévenin equivalent output impedance, equation 83
φ_A = flux of unit A, equations 17 and 18
φ_B = flux of unit B, equations 17 and 18
φ_G = gate flux per leg, either left or right
$\varphi_M = 10^8\ E_{XM}/4N_X\omega$
φ_S = saturation flux per leg, equation 26
ψ_C = control circuit flux linkages, equation 15

APPROACH

The analysis will be directed toward applying and extending the methods of Storm[1-3] to include the actions of the linear and partially nonlinear circuit into which the saturable reactor unit is connected. The fundamental analyses which Storm has produced serve as the basic background and as an analytical foundation without which it would have been impractical to undertake this work. In the circuit under consideration the application of direct current produces saturation in the saturable-reactor unit unbalancing the voltage division between units resulting in output. In this device with low control-circuit impedance, two modes of operation exist; one when the unit without signal acts as a transformer at all times, and another mode when the alternating voltages are sufficient to saturate the unit without signal. Assuming a core

---

Paper 57-96, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 17, 1956; made available for printing December 11, 1956.

R. E. ANDERSON is with the General Electric Company, Schenectady, N. Y.

The author wishes to express his thanks to Dr. H. F. Storm and H. L. Kellogg of the General Electric Company for their advice and support in the preparation of this paper.

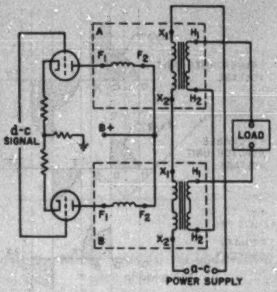


Fig. 1. Basic magnetic modulator circuit employing saturable-core transformers

material having a rectangular dynamic hysteresis loop[4] as shown in Fig. 3, the various flux paths will be either saturated or unsaturated. On this basis a point of saturation, α, can be expressed in terms of electrical angle in time. The analysis will proceed by writing equations for the signal and output in terms of firing angle α. These equations will be solved to obtain the current gain, then such results will form the basis of performance graphs and expressions to be used in application. The highlights of this analysis will be shown in the following paragraphs while the details may be found in the appendix. The equation numbers given in the following refer to the corresponding equations in the appendix.

The gain in mode 1, transformer unsaturated, will be evaluated by considering that portion of the cycle previous to d-c saturation $0<\omega t<\alpha$ called the exciting interval. Then, neglecting the exciting current drop, the line voltage divides between the two units equally and

$$2E_{BX}=\frac{1}{\pi}\int_0^{\alpha/\omega} E_{XM}\sin\omega t\,dt \quad (0<\omega t<\alpha) \tag{11}$$

$$=E_{XM}\times\frac{(1-\cos\alpha)}{\pi} \quad (0<\omega t<\alpha) \tag{12B}$$

After the saturable reactor saturates, $\omega t>\alpha$, the applied voltage drives the load and the output becomes

$$I_H=\frac{\displaystyle\int_{\alpha/\omega}^{\pi/\omega} E_{MX}\sin\omega t}{R_O\left(\frac{N_X}{N_H}\right)^2+R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]} \quad (\alpha<\omega t<\pi) \tag{42}$$

By expressing the net gate-circuit magnetomotive force in terms of primary current the signal requirements for this output are given by the following equation

$$N_CI_C=\left[\frac{2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2}{1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2}\right] N_X\frac{\displaystyle\int_{\alpha/\omega}^{\pi/\omega} i_X\,dt}{\pi/\omega-0} \tag{40}$$

Eliminating α between these equations by suitable mathematics yields the mode 1 gain equation of

$$N_HI_H=\frac{N_CI_C}{\left[2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]} \tag{44}$$

When the applied line voltage is sufficiently high the transformer unit may undergo a-c saturation. Since the transformer voltage changes with signal, this condition may exist for some signal regions and may not exist for other signals. If the applied voltage is so great that a-c saturation of both units occurs without signal the primary current is limited by primary resistance alone.

$$I_X=\frac{E_X-E_{SX}}{R_X} \quad \begin{matrix}(\mathfrak{X}>1)\\(I_C=0)\end{matrix} \tag{49}$$

Under these overexcited conditions the a-c magnetomotive force establishes a minimum signal level below which no controlling action can take place. This establishes a minimum signal level of

$$N_CI_{C\,\min}=N_X\frac{E_X-E_{SX}}{R_X} \quad (\mathfrak{X}>1) \tag{50}$$

Denoting the transformer saturation instant in electrical time angle as β, the relationship between this angle and the saturable reactor firing angle, α, may be obtained from a knowledge of the transformer saturation voltage or saturation flux density as

$$\frac{\displaystyle\int_0^{\alpha/\omega}\frac{E_{XM}}{2}\sin\omega t\,dt+\int_{\alpha/\omega}^{\beta/\omega}\times(E_{XM}\sin\omega t-R_X)dt}{\pi/\omega-0}=\frac{E_{SX}}{2} \tag{51}$$

To obtain the gain equation in mode 2 it will be necessary to eliminate two parametric variables, α and β. The voltage equation will serve as the additional equation in these unknowns. In a manner similar to the mode 1 case the equation for the output current will be written. As has been mentioned previously, the after-saturation impedance is assumed negligible and there is no output from β to π. The theoretical wave shapes have been illustrated in Fig. 4 to aid the reader in visualizing the circuit operation and the mathematical limits. Thus during the mode 2 operation the output current equation will involve integration limits of α to β providing

$$I_H=\frac{\displaystyle\int_{\alpha/\omega}^{\beta/\omega}\frac{E_{HM}\sin\omega t}{R_O\left(1+\frac{R_X}{R_M}\right)+R_B\left(\frac{N_H}{N_X}\right)^2}dt}{(\pi/\omega-0)} \tag{55}$$

Now the signal requirements may be ex-

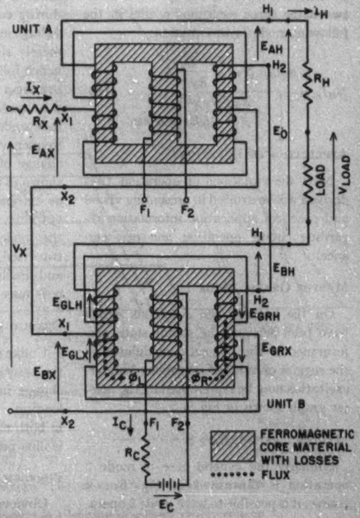


Fig. 2. Complete circuit diagram showing polarities and direction of current flow

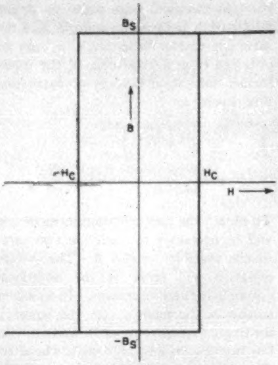


Fig. 3. Rectangular hysteresis loop

pressed in terms of the firing and saturation angles hence by circuit analysis as shown in the appendix the control equation becomes

$$N_CI_C=N_X\left[\frac{\displaystyle\int_{\alpha/\omega}^{\beta/\omega}\frac{E_{XM}}{R_X+\dfrac{R_M\,R_O\left(\dfrac{N_X}{N_H}\right)^2}{R_M+R_O\left(\dfrac{N_X}{N_H}\right)^2}}\times\left[\frac{2+\dfrac{R_O}{R_M}\left(\dfrac{N_X}{N_H}\right)^2}{1+\dfrac{R_O}{R_M}\left(\dfrac{N_X}{N_H}\right)^2}\right]\sin\omega t\,dt+\int_{\beta/\omega}^{\pi/\omega}\frac{E_{XM}}{R_X}\sin\omega t\,dt}{\pi/\omega-0}\right] \tag{53}$$

Mathematically eliminating α and β between the three equations results in the following mode 2 gain equation:

$$N_HI_H=\frac{N_CI_C-N_X\dfrac{(E_X-E_{SX})}{R_X}}{\left[1+\dfrac{R_O}{R_X}\left(\dfrac{N_X}{N_H}\right)^2\right]} \tag{59}$$

APPLICATION OF EQUATIONS

Next the equations of operation thus derived will be utilized in presenting visual and practical application information regarding circuit operation and performance.

MASTER GRAPH, MODE 1

On the basis of the equations which have been derived, the steady-state performance of the magnetic modulator in the regions of one-half normal and underexcitation may be represented by the master graph shown in Fig. 5.

MASTER GRAPH, MODE 2

Similarly the general case of mode 2 operation is illustrated in Fig. 6. Of course, it is possible to have mode 2 operation at all signal levels rather than the general transition case shown. Probably the greatest fundamental difference between mode 1 and mode 2 operation is revealed in the current-gain equations. Here it is found that the ratio of load-circuit resistance to magnetizing resistance, R_0/R_M, controls the gain in mode 1 while in mode 2 the ratio of load-circuit resistance to primary-circuit resistance, R_0/R_X, is the controlling factor.

Load Current Versus Supply Voltage and Frequency

From the equations of operation which have been derived in the analysis it is possible to construct the functional relationship between load current and primary voltage. This performance, illustrated by Fig. 7, consists of three regions of operation. With very low line voltage the output will be resistance limited; at intermediate voltages the output will be mode 1 signal- or mode 2 saturation-controlled. When the line voltage is increased into the signal-controlled area of mode 2 operation the output will drop during constant signal operation as line voltage increases. The effect of frequency merely shifts the saturation voltage and hence follows directly from Fig. 7. (Assuming a rectangular hysteresis loop with a unique saturating flux density that is independent of frequency, the effect of frequency change may be viewed as a change of saturation voltage, E_{SX}.) As shown in the frequency diagram of Fig. 8, at extremely low frequencies there is no output. With increasing frequency first the mode 2 saturation-limited region is traversed, then mode 2 signal-limited, and finally at high frequencies mode 1 or resistance-limited operation ensues.

OUTPUT VERSUS LOAD

Unlike the ordinary saturable reactor,[1,2] the saturable-core transformer exhibits a change in current gain with changes in loading. This condition, which is found in both mode 1 and in mode 2 operation, is illustrated by Fig. 9.

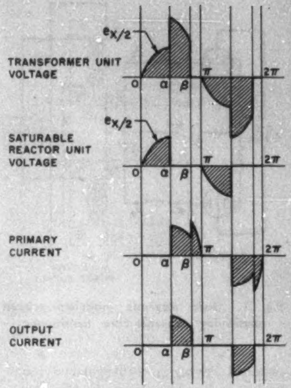


Fig. 4. Theoretical modulator wave shapes during mode 2 operation

PERFORMANCE EQUATIONS

Obviously, the output power delivered to the load represents the transformer rating. Thus, the load power is given by

$$P_{L\text{rms}}=I_M{}^2R_{\text{load}}\times k_f{}^2 \tag{88}$$

After defining the matching index, n, it is possible to write the expression for maximum power output which occurs when $\alpha=0$,

$$P_{L\max}=\frac{E_H{}^2\,k_f{}^2}{4\left(1+\dfrac{R_X}{R_M}\right)^2R_{TH}}\times\frac{4n}{(n+1)^2} \tag{85}$$

where

$$n=\frac{R_{\text{load}}}{R_{TH}} \tag{84}$$

and the output impedance, R_{TH},

$$R_{TH}=R_H+\frac{R_MR_X\left(\dfrac{N_H}{N_X}\right)^2}{R_M+R_X} \tag{83}$$

Applying the maximum power transfer theorem at full output, $\alpha=0$, produces the obvious result that $n=1$ for maximum power transfer. From the previous derivations the power gain may be written as

$$\text{Mode 1 } K_{P1}=n\,\frac{R_{TH}}{R_C'}\,k_f{}^2\times\frac{1}{\left[2+\dfrac{R_O}{R_M}\left(\dfrac{N_X}{N_H}\right)^2\right]^2} \tag{86}$$

and

$$\text{Mode 2 } K_{P2}=n\,\frac{R_{TH}}{R_C'}\,k_f{}^2\times\frac{1}{\left[1+\dfrac{R_O}{R_X}\left(\dfrac{N_X}{N_H}\right)^2\right]^2} \tag{87}$$

The current gains are given the approach section as the following equations

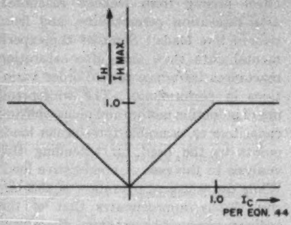


Fig. 5. Master graph for the underexcited region of operation

$$K_{I_1}=\frac{N_C/N_H}{\left[2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]}\qquad(\alpha<0°)\qquad(80B)$$

and

$$K_{I_2}=\frac{N_C/N_H}{\left[1+\frac{R_O}{R_X}\left(\frac{N_X}{N_H}\right)^2\right]}\qquad(\alpha<0°)\qquad(82B)$$

From these equations the voltage gain may be written as

$$K_{V_1}=\frac{nN_CR_{TH}}{N_HR_C\left[2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]}\qquad(\alpha<0°)\qquad(80C)$$

and

$$K_{V_2}=\frac{nN_CR_{TH}}{N_HR_C\left[1+\frac{R_O}{R_X}\left(\frac{N_X}{N_H}\right)^2\right]}\qquad(82C)$$

Experimental Correlation

For the purpose of experimentally verifying this analysis, and for determining the magnitude of certain limitations introduced by the assumptions of the analysis, an experimental transformer was constructed. In addition to the functions just mentioned this activity served as an introduction to many of the practical problems associated with the saturable-core transformer.

EXPERIMENTAL TRANSFORMER

The experimental transformer employed in this work was designed especially for laboratory study by using a 4-core construction with multiple leads and taps to facilitate the study of all flux conditions. The use of an oversize model permitted ease of measurement, while a toroidal construction reduced the exciting current and leakage reactance. One-mil-thick core material of Deltamax was chosen for its close representation of the assumed rectangular hysteresis loop. Such a material under sinusoidal flux conditions exhibits a square-wave exciting current having the fundamental component in phase with the exciting voltage. Thus it comes quite close to the case of the ideal rectangular hysteresis loop, where there would be no reversible energy storage and the magnetizing branch would be a resistance.

MEASURING TECHNIQUE

Since the theory relates the circuit parameters to the average values of voltage and current, a rectifier-type voltmeter was employed to measure circuit voltages. Resistive meter shunts permitted similar a-c measurements while d'Arsonval ammeters indicated the d-c component in the signal circuits. The extremely low core loss of Deltamax necessitated the use of special light-beam wattmeters for the evaluation of core loss.

EXPERIMENTAL RESULTS

The experimental magnetic modulator circuit was operated under various conditions of load, voltage, frequency, and signal. In addition, the transformer parameters were varied externally to represent different design conditions. A representative example of the many output versus signal conditions studied is given in Fig. 10, which shows mode 1 operation. In this curve, as in all subsequent curves, the experimentally measured points are indicated by circles while the theoretical calculations are shown by solid lines. The observed fact that the maximum output obtained experimentally did not quite reach the theoretical value may be the consequence of transformer leakage reactance drop. Fig. 11 includes both mode 1 and mode 2 operations, where the high values in mode 2 operation are also believed to originate from leakage

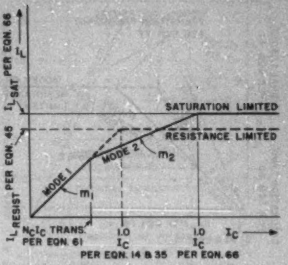
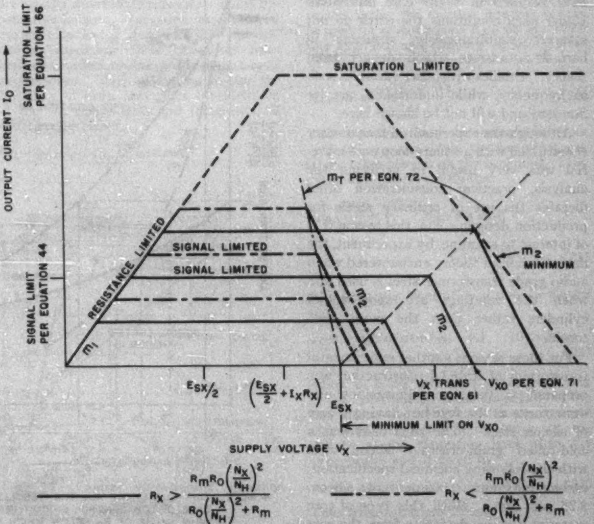


Fig. 6. Master graph for the region of operation where $X>1/2$



GRAPH LINES REPRESENT LOCI OF CONSTANT SIGNAL OPERATION

Fig. 7. Output current as a function of primary supply voltage

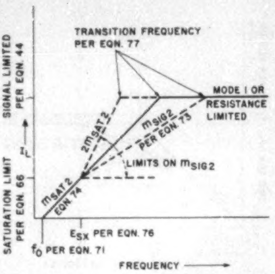


Fig. 8. Effect of frequency on modulator output

reactance conditions, particularly in the primary circuit.

The effect of line-voltage variation on output has been given thorough consideration in the analyses. Since this consideration is of major importance in military applications, where the voltage range is notoriously wide, two experimental studies, Fig. 12 and Fig. 13, are presented. Fig. 12 is an example where the maximum output is saturation-limited, while Fig. 13 is an example of signal-limited maximum. In both these examples the output is never completely reduced to zero by high overvoltage since manufacturing variations have resulted in slight core mismatch. Under such conditions the cores do not saturate simultaneously, resulting in harmonic outputs which depart from theory. Studies with other variables such as frequency, while undertaken, are redundant and will not be shown here.

Although the experimental transformer constructed with a square-loop core material was very useful in verifying the analysis, practical consideration often dictates the use of ordinary steels for production designs. For this reason it is of interest to examine, by experiment, the deviations from theory encountered when audio-grade transformer steel is used and when the windings are spool-wound cylinders rather than the more ideal toroid.

For these reasons another experimental transformer of toroidal construction was prepared to the same dimensions and same turns as the first but having a core of silicon steel. (The steel used was a cold-rolled grain-oriented silicon steel with the following chemical specification: carbon, maximum, 0.006 per cent; silicon, 3.25 per cent.) With this type of core material a major deviation from the core characteristic assumed for analysis was found in the after-saturation permeability or slope to the top of the *B-H* loop. Under these conditions the performance in mode 2 operation began to depart from theory to varying extents as shown in Figs. 11 and 13. However, when the 3-legged construction is employed the resulting characteristics lack the clear-cut corners found in the performance of the toroidal device. An example of the 3-legged construction was prepared using two cut *C* cores per reactor. The performance of this construction is illustrated by Figs. 14 and 15.

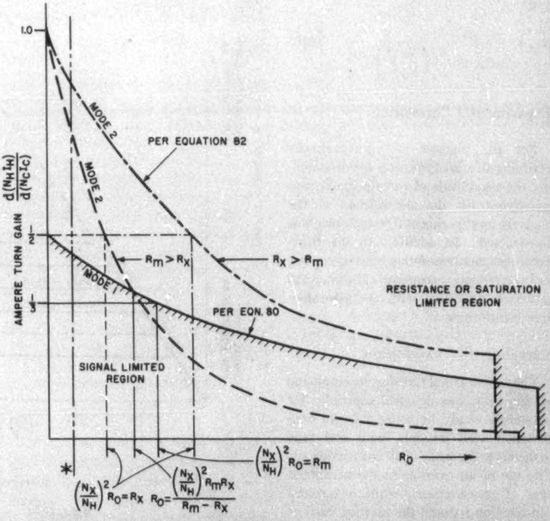
Conclusions and Future Work

The steady-state operation of the saturable-core transformer has been analyzed as a magnetic amplifier with resistive load. This analysis reveals the nature of operation and explains the important influence of core loss on the amplifier gain. The simple, concise equations resulting from this analysis simplify calculations of performance as well as serving as a sound basis for design. Experimental results have been presented to show the trends encountered in practical devices which approach but do not fully meet the assumptions used to simplify analysis.

Nevertheless, this analysis ignored the effect arising from leakage reactance, after-saturation permeability, and from nonresistive loads. Some of the experimental data show that after-saturation reactances introduce second-order variations in performance. The widespread use of induction motors and phase-shifting capacitors as saturable transformer loads points to the need for extending this analysis to the case of nonresistive load. While the transient operation of this circuit closely approximates that of the ordinary saturable reactor, there are experimentally observed transient conditions which could bear further investigation.

Appendix

The circuit to be analyzed is shown in Fig. 2. For purposes of analysis the signal is applied to one unit, the *B* unit, which acts as a saturable reactor while the other unit, unit *A*, performs as a transformer and as a load for unit *B*. Although in practice push-pull signals may be applied to windings of both units, it may be shown experimentally that the signal on a single unit is equivalent to the differential push-pull signal. The symbols for the various circuit voltages and currents together with the polarity directions used in analysis are shown in Fig. 2.



* NOTE: THIS CURVE, LIKE ALL THE RESULTS OF THIS ANALYSIS, IS BASED ON THE ASSUMED CASE OF LOW CONTROL CIRCUIT IMPEDANCE. THIS ASSUMPTION ESTABLISHES A MINIMUM VALUE OF R_O BELOW WHICH THE ABOVE CURVES ARE NO LONGER VALID.

Fig. 9. Current gain as a function of loading

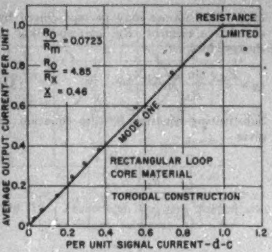


Fig. 10. Control characteristic underexcited

Assuming a low control-circuit impedance, $R_C' << R_O$, the gate windings of any given unit will be coupled through the control winding during the unsaturated intervals. When saturation occurs in one leg of the reactor, the low control-circuit impedance will be reflected as a short circuit in the unsaturated circuit. Hence the following result.

VOLTAGE EQUATIONS

$$E_{GL} = E_{GR} = E_G \tag{1}$$

and

$$E_{BX} = 2E_{BGX} \tag{2}$$

Applying Kirchhoff's potential law to the primary and to the output circuits gives

$$V_X = V_{AX} + V_{BX} = E_{AX} + E_{BX} + I_X R_X \tag{3}$$

$$E = {}_O E_{AH} - E_{BH} \tag{4}$$

Multiplying equation 3 by the *X-H* turns ratio and substituting into equation 4 produces

$$E_O = (V_X - I_X R_X)\frac{N_H}{N_X} - 2E_{BH} \tag{5}$$

Substituting equation 2 into equation 5 gives

$$E_O = (V_X - I_X R_X)\frac{N_H}{N_X} - 4E_{GH} \tag{6}$$

where

$$V_X - I_X R_X = E_X \tag{7}$$

If during the unsaturated period, known as the exciting interval, the incremental permeability is unique and single-valued, the incremental inductance of unit *A* and unit *B* will be equal. From a practical viewpoint this assumption implies that the sides of all hysteresis loops be parallel. Then the induced portion of the applied voltage, E_X, will divide equally between the two units as

$$e_{AX} = e_{BX} \text{ or } e_{AH} = e_{BH} \tag{8}$$

On substituting equation 8 into equation 4 it is seen that during the exciting interval

$$e_O = 0 \qquad (0 < \omega t < \alpha) \tag{9}$$

From equations 6 and 7 the output voltage in steady state is

$$E_O = E_X - 4E_{GH} \tag{10}$$

Mode 1

Assuming an ideal core material with negligible after-saturation inductance, E_G will be zero during the saturation interval. Thus, the output voltage may be obtained from equation 10 with a knowledge of E_G during the exciting interval. Since the applied voltage is equally divided between the two units during the exciting interval, the gate voltage may be found by integration as follows

$$4E_{GH} = 2E_{BH} = \frac{1}{\pi}\int_0^{\alpha/\omega} E_{HM} \sin \omega t \, dt \qquad (0 < \omega t < \alpha) \tag{11}$$

$$= \frac{E_{HM}}{\pi}(1 - \cos \alpha) \tag{12}$$

Substituting equation 12 into equation 10 results in

$$E_O = \frac{(1 + \cos \alpha)}{2} E_H \tag{13}$$

FLUX RELATIONSHIPS

In view of the need for considering the flux conditions when designing, the various flux relationships will be analyzed. This analysis will proceed by writing each of the equations for the control, power, and output circuits using Kirchhoff's potential law and Faraday's induced voltage law. Thus for the control circuit,

$$E_C - \frac{d\varphi_L}{dt} \times \frac{N_C}{10^8} + \frac{d\varphi_R}{dt} \times \frac{N_C}{10^8} - i_C R_C = 0 \tag{14}$$

For purpose of analysis the control-circuit impedance has been assumed resistive. Although in practice the control windings are usually shunted by capacitors which bypass the induced harmonics, experimental evidence indicates that with low control-circuit impedance these capacitors have little or no effect on the steady-state circuit performance. Thus neglecting R_C for the free flow of even harmonics consideration of the a-c components and integration of equation 14 yields

$$(\varphi_L - \varphi_R) N_C = \psi_C, \text{ a constant} \tag{15}$$

or

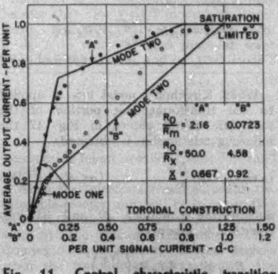


Fig. 11. Control characteristic transition excitation

A—Audio-grade-transformer steel-core material

B—Rectangular-loop-core material

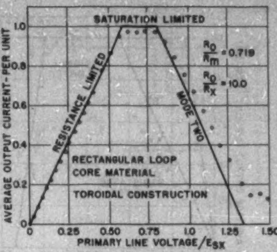


Fig. 12. Influence of supply voltage on output-saturation-limited case

$$\frac{d\varphi_L}{dt} = \frac{d\varphi_R}{dt} \tag{16}$$

Similarly, one may write the flux equation for the power and load circuits as

$$e_X - \frac{d\varphi_A}{dt}\frac{N_X}{10^8} - \frac{d\varphi_B}{dt}\frac{N_X}{10^8} - i_X R_X = 0 \tag{17}$$

and

$$\frac{d\varphi_A}{dt}\frac{N_H}{10^8} - \frac{d\varphi_B}{dt}\frac{N_H}{10^8} - i_H R_O = 0 \tag{18}$$

where

$$R_O = R_{HA} + R_{HB} + R_{\text{load}} \tag{19}$$

During the exciting interval the *X* winding current is limited to the exciting current which may be considered small. For this reason during the exciting interval v_x may be used for e_x and equals $E_{XM} \sin \omega t$. With this approach and defining φ_B as equal to $\varphi_L + \varphi_R$, equation 17 may be solved as

$$E_{AM} \sin \omega t - 4\frac{d\varphi_G}{dt}\frac{N_X}{10^8} = 0 \tag{20}$$

This equation may be integrated from $t = 0$ to α/ω to obtain the flux swing $\varphi_{G\Delta}$ which takes place from the beginning of the cycle until the core saturates.

$$4\varphi_{G\Delta}\frac{N_X}{10^8} = \int_0^{\alpha/\omega} E_{XM} \sin \omega t \, dt$$

$$= \frac{(1 - \cos \alpha)}{4f} E_X \tag{21}$$

Also from this integration it may be shown that

$$\varphi_L = -\varphi_M \cos \omega t + [\varphi_m + \varphi_1] \qquad (0 < \omega t < \alpha) \tag{22}$$

where φ_1 is the flux level at the beginning of the cycle and φ_m is the magnitude of the flux alternating component in steady-state operation. The flux in the other leg, φ_R, will be at the saturation level when the cycle begins and is given by

$$\varphi_R = -\varphi_m \cos \omega t \; \varphi_m - \varphi_S \tag{23}$$

Thus from equations 22 and 23

$$\varphi_L - \varphi_R = \varphi_1 + \varphi_S \tag{24}$$

Accordingly it is seen that the fluxes in

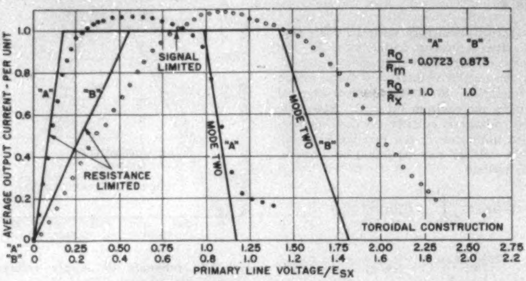


Fig. 13. Influence of supply voltage on output-signal-limited case

A—Rectanguler-loop-core material
B—Audio-grade-transformer steel-core material

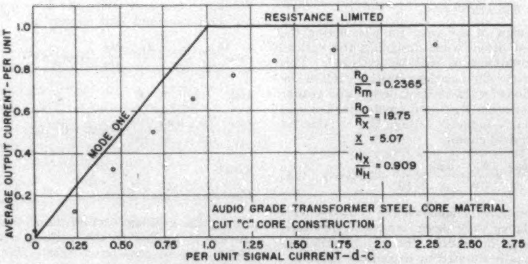


Fig. 14. Control characteristic underexcited

unit B vary in a manner identical to the fluxes found in the ordinary saturable reactor as described by Storm;[1] see Fig. 16. From geometry, as pointed out by Storm,

$$\varphi_1 = \varphi_S - \varphi_G \Delta \tag{25}$$

If the saturation voltage of the pair of transformers is defined by

$$\varphi_S = \frac{1}{16f} E_{SX} \frac{10^8}{N_X} \tag{26}$$

The initial flux φ_1, as well as the flux difference $\varphi_L - \varphi_R$ may be expressed in terms of the line voltage and the firing angle. Thus

$$\varphi_1 = \frac{10^8}{16fN_X} [E_{SX} - (1 - \cos\alpha) E_X] \tag{27}$$

and

$$\varphi_L - \varphi_R = \frac{10^8}{16fN_X} [2E_{SX} - (1 - \cos\alpha) E_X] \tag{28}$$

CURRENT EQUATIONS

With these flux and flux voltage equations as a basis, expressions for the circuit currents will be derived. For this purpose the simplified circuit of Fig. 17 will be used. Neglecting the after-saturation inductance, during the conduction or saturated interval,

$$\frac{d\varphi_B}{dt} = 0 \text{ and hence } e_B = 0 \tag{29}$$

which together with Kirchhoff's and Faraday's laws applied to the output circuit of Fig. 17 may be substituted into equation 17 to yield

$$V_{MX} \sin\omega t - (i_H R_O) \frac{N_X}{N_H} - i_X R_X = 0 \qquad (\alpha < \omega t < \pi) \tag{30}$$

When Kirchhoff's nodal law is applied to the transformer-unit portion of the schematic circuit shown in Fig. 17 the following relationship is obtained:

$$i_H = (i_X - i_M) \frac{N_X}{N_H} \qquad (\alpha < \omega t < \pi) \tag{31}$$

Substituting equations 29 and 31 into equation 30 produces

$$i_X = \frac{E_{MX} \sin\omega t + i_M R_O \left(\frac{N_X}{N_H}\right)^2}{\left[R_O \left(\frac{N_X}{N_H}\right)^2 + R_X\right]} \qquad (\alpha < \omega t < \pi) \tag{32}$$

Assuming that the magnetizing branch of the transformer may be represented by a resistive element, R_M, corresponding to the core loss, let

$$i_M = \frac{E_{MX} \sin\omega t - i_X R_X}{R_M} \tag{33}$$

Substituting equation 33 into equation 32 gives

$$I_X = \frac{\left[1 + \frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]}{\left(\frac{N_X}{N_H}\right)^2 R_O + R_X\left[1 + \frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]} \times \frac{\int_{\alpha/\omega}^{\pi/\omega} E_{MX} \sin\omega t \, dt}{\pi/\omega - 0} \qquad (\alpha < \omega t < \pi) \tag{34}$$

which by integration becomes

$$I_X = \frac{\left[1 + \frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right] E_{MX}}{R_O\left(\frac{N_X}{N_H}\right)^2 + R_X\left[1 + \frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]} \times \frac{(1 + \cos\alpha)}{\pi} \qquad (\alpha < \omega t < \pi) \tag{35}$$

CONTROL EQUATIONS

To introduce the control function of the signal windings, the magnetomotive force equation for the control magnetic path (leg) may be written as follows:

$$N_X i_X + N_H i_H - N_C i_C = 0 \tag{36}$$

This magnetomotive force equation is equated to zero because there is always one flux path, φ_L or φ_R, which is unsaturated and because the width of the B-H loop may be neglected in comparison to the controlling and gating magnetomotive forces. During the exciting interval, $i_H = 0$ and the control equation show that with neither flux path saturated,

$$N_C i_C = N_X i_X = N_X i_{XL} - N_X i_{XR} = 0 \qquad (0 < \omega t < \alpha) \tag{37}$$

During the saturation interval the simplified transformer circuit of Fig. 18 may be considered as a linear load on the saturable-reactor unit. Applying Kirchhoff's laws

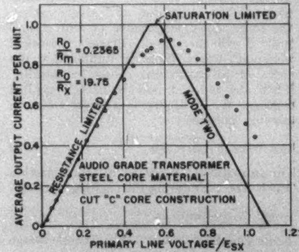


Fig. 15. Influence of supply voltage on output-saturation-limited case

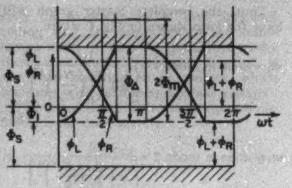


Fig. 16. Theoretical flux wave shapes (after Storm, reference 1)

to this circuit shows that

$$I_H=\frac{1}{\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2}\left(\frac{N_X}{N_H}\right)I_X \tag{38}$$

Thus, from equations 36 and 38,

$$N_Ci_C=N_Xi_X+N_Hi_H$$
$$=\left[\frac{2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2}{1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2}\right]N_Xi_X \quad (\alpha<\omega t<\pi) \tag{39}$$

Using equations 37 and 39 the control current may be related to the primary current by integration,

$$N_Ci_C=\int_0^{\alpha/\omega}\frac{N_Xi_Xdt}{\pi/\omega-0}+\int_{\alpha/\omega}^{\pi/\omega}\frac{(N_Xi_X+N_Hi_H)}{\pi/\omega-0}dt$$
$$=\left[\frac{2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2}{1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2}\right]N_X\int_{\alpha/\omega}^{\pi/\omega}\frac{i_Xdt}{\pi/\omega-0} \tag{40}$$

Substituting equation 35 into equation 40 yields

$$N_Ci_C=\frac{N_X\left[2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]E_{MX}}{R_O\left(\frac{N_X}{N_H}\right)^2R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]}\times\frac{(1+\cos\alpha)}{\pi} \tag{41}$$

LOAD CURRENT

From Fig. 18 the output current during the saturation interval is seen to be

$$I_H'=\frac{\int_{\alpha/\omega}^{\pi/\omega}E_{MX}\sin\omega t}{R_O\left(\frac{N_X}{N_H}\right)^2+R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]} \quad (\alpha<\omega t<\pi) \tag{42}$$

Since the output-circuit current is zero during the exciting interval, the total output current may be obtained directly from the integration of equation 42 as

$$I_H'=\frac{N_HI_H}{N_X}=\frac{(1+\cos\alpha)E_{MX}}{\left\{R_O\left(\frac{N_X}{N_H}\right)^2+R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]\right\}\pi} \tag{43}$$

Substituting equation 42 into equation 43 and multiplying by the $X-H$ turns ratio produces

$$N_HI_H=\frac{N_CI_C}{\left[2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]} \tag{44}$$

The maximum output may be evaluated by letting $\alpha=0$ degrees. Thus

$$I_{H\ \max}=\frac{E_H}{R_O\left(1+\frac{R_X}{R_M}\right)+R_X\left(\frac{N_H}{N_X}\right)^2} \quad (\alpha=0°) \tag{45}$$

This is known as the resistance-limited case, and may be rearranged to show the matching term and a Q term as

$$I_{H\max}=\frac{E_H}{R_O}\,\frac{1}{\left[\left(1+\frac{R_X}{R_M}\right)+\frac{R_X}{R_O}\left(\frac{N_H}{N_X}\right)^2\right]} \tag{46}$$

From these equations, Fig. 5 follows.

Mode 2

EXCITATION LEVEL

In the manner explained previously the excitation level establishes the mode of operation. For purposes of analysis the per-unit excitation, $\mathcal{X}$, is defined by equation 47 as

$$\mathcal{X}=\frac{E_X}{E_{SX}} \tag{47}$$

With this definition the saturable-reactor unit will have normal excitation when $\mathcal{X}=1$. On the other hand, the transformer receives from half to full line voltage depending upon the firing angle; a condition which requires $\mathcal{X}<1/2$ to avoid a-c saturation of the transformer. Furthermore with $\mathcal{X}>1$ a-c saturation of the transformer occurs with no signal. Thus, since mode 2 operation is characterized by the saturation of the transformer unit, it may be seen that, for $\mathcal{X}<1/2$ mode 1 operation exists, for $1/2<\mathcal{X}<1$ a transition from mode 1 to mode 2 takes place at some minimum signal level, and for $\mathcal{X}>1$ mode 2 operation is found to exist at least for small signals.

When the transformer unit saturates the saturable-reactor unit will be in saturation and the primary current will be limited only by the primary resistance. Thus defining β as the time in electrical degrees when the transformer unit saturates, the primary current during the latter part of the half-cycle is obtained by applying Ohm's law to the circuit of Fig. 2. With the previous assumption of zero after saturation core permeability, the following is obtained:

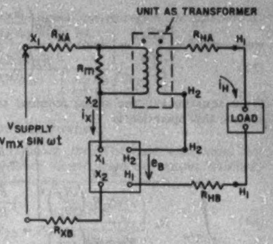


Fig. 17. Simplified schematic diagram for analysis of magnetic modulator circuit

$$i_X=\frac{E_{MX}\sin\omega t}{R_X} \quad (\beta<\omega t<\pi) \tag{48}$$

Provided the after-saturation inductances of the two units are equal, during this overexcitation portion of the cycle the output voltage will be zero. When $\mathcal{X}>1$ and with no applied signal, both units will saturate, thus permitting a primary current to flow as described by equation 48 and having an average value of

$$I_X=\frac{E_X-E_{SX}}{R_X} \quad \begin{matrix}(\mathcal{X}>1)\\(I_C=0)\end{matrix} \tag{49}$$

Control signals which are too small to limit the core saturation to one polarity will have no controlling effect on the saturable-reactor unit. Thus the minimum control signal will have to supply d-c ampere-turns equal to or greater than the a-c exciting ampere-turns. This minimum signal may be evaluated by substituting equation 49 into the integral of equation 36 with $i_H=0$, as

$$N_CI_{C\ \min}=N_X\frac{E_X-E_{SX}}{R_X} \quad (\mathcal{X}>1) \tag{50}$$

MODE 2 EQUATIONS

The wave shapes shown in Fig. 4 illustrate the nature of mode 2 operation and are utilized to determine the mathematical limits in the equations.

Thus, the analysis may be undertaken by using the transformer a-c saturation equation for obtaining the relation between α and β.

$$\frac{E_{SX}}{2}=\frac{\int_0^{\alpha/\omega}\frac{E_{XM}}{2}\sin\omega t\,dt+\int_{\alpha/\omega}^{\beta/\omega}(E_{XM}\sin\omega t-i_XR_X)dt}{\pi/\omega-0} \tag{51}$$

Integrating and rearranging equation 51

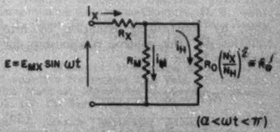

I_X
R_X
$E=E_{MX}$ SIN ωt
R_M
i_M
i_H
$R_O\left(\frac{N_X}{N_H}\right)^2=R_O'$
$(\alpha<\omega t<\pi)$

Fig. 18. Simplified load circuit figure

$$\cos\alpha-\cos\beta=\left[\frac{2E_{SX}}{E_X}-1+\cos\beta\right]\frac{R_O\left(\frac{N_X}{N_H}\right)^2+R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]}{R_O\left(\frac{N_X}{N_H}\right)^2-R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]} \tag{52}$$

From equation 40, the signal required to achieve this operation is

$$N_CI_C=\frac{\displaystyle\int_{\alpha/\omega}^{\beta/\omega}\frac{E_{XM}}{R_X+\dfrac{R_MR_O\left(\frac{N_X}{N_H}\right)^2}{R_M+R_O\left(\frac{N_X}{N_H}\right)^2}}\left[\frac{2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2}{1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2}\right]\sin\omega t\,dt+\int_{\beta/\omega}^{\pi/\omega}\frac{E_{XM}}{R_X}\sin\omega t\,dt}{\pi/\omega-0} \tag{53}$$

and by integration

$$\left[\frac{2N_CI_C}{N_XE_X}-\frac{1}{R_X}\right]=\frac{\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right](\cos\alpha-\cos\beta)}{\left\{R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]+R_O\left(\frac{N_X}{N_H}\right)^2\right\}\left[\dfrac{1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2}{2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2}\right]}+\frac{\cos\beta}{R_X} \tag{54}$$

Next the output current will be evaluated under these conditions using equation 42 together with the appropriate limits of α and β.

$$I_H=\frac{\displaystyle\int_{\alpha/\omega}^{\beta/\omega}\frac{E_{HM}\sin\omega t}{R_O\left(1+\frac{R_X}{R_M}\right)+R_X\left(\frac{N_H}{N_X}\right)^2}dt}{(\pi/\omega=0)} \tag{55}$$

From the integration of equation 55

$$I_H=\frac{E_X\left(\frac{N_X}{N_H}\right)(\cos\alpha-\cos\beta)}{2\left\{R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]+R_O\left(\frac{N_X}{N_H}\right)^2\right\}} \tag{56}$$

Substituting equation 54 into equation 56 yields

$$I_H=\frac{N_CI_C-\frac{E_XI_X}{2R_X}\;(1+\cos\beta)}{N_H\left[2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]}. \tag{57}$$

Next the term $(1+\cos\beta)$ is evaluated by substituting equation 52 into equation 54 resulting in

$$\frac{(1+\cos\beta)}{2}=\frac{\left\{R_O\left(\frac{N_X}{N_H}\right)^2-R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]\right\}\frac{N_CI_C}{N_XE_X}-\left[2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]\left[\frac{E_{SX}}{E_X}-1\right]}{2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2+\frac{1}{R_X}\left\{R_O\left(\frac{N_X}{N_H}\right)^2-R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]\right\}} \tag{58}$$

Now, by substituting equation 58 into equation 57, the control relationship for mode 2 is found to be

$$N_HI_H=\frac{N_CI_C-N_X\frac{(E_X-E_{SX})}{R_X}}{\left[1+\frac{R_O}{R_X}\left(\frac{N_X}{N_H}\right)^2\right]} \tag{59}$$

MASTER GRAPH

From the foregoing derivation the master graph including mode 2 may be prepared as shown in Fig. 6. For this purpose the following equations have been obtained by rearranging the foregoing general equations: The boundary for mode 2 operation is, $X>1/2$ and

$$E_X>\frac{E_{SX}}{1+\frac{(1+\cos\alpha)}{2}\times\dfrac{R_O\left(\frac{N_X}{N_H}\right)^2-R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]}{R_O\left(\frac{N_X}{N_H}\right)^2+R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]}} \tag{60}$$

Transition from mode 1 to mode 2 takes place when

$$N_CI_{C\,\mathrm{trans}}=\frac{N_X\,\frac{E_{SX}-E_X}{R_X}}{\left[\dfrac{1+\frac{R_O}{R_X}\left(\frac{N_X}{N_H}\right)^2}{2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2}\right]-1}\qquad(E_X>E_{SX}) \tag{61}$$

The current gain in mode 1 and mode 2 are equal when

$$R_X=\frac{R_MR_O\left(\frac{N_X}{N_H}\right)^2}{R_O\left(\frac{N_X}{N_H}\right)^2+R_M} \tag{62}$$

In this case, mode 2 operation begins when $E_X\geq E_{SX}$, regardless of signal.

Thus, the complete master graph will have the form shown in Fig. 6 where

$$m_1=\text{slope in mode 1}=\frac{1}{\left[2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]} \tag{63}$$

$$m_2=\text{slope in mode 2}=\frac{1}{\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]} \tag{64}$$

The maximum output in the resistance-limited case is determined from equation 45,

$$I_{H\,\max}=\frac{E_H}{R_O\left(1+\frac{R_X}{R_M}\right)+R_X\left(\frac{N_H}{N_X}\right)^2}\quad\begin{array}{l}(\alpha=0^\circ)\\(\beta<\pi)\end{array} \tag{65}$$

While the maximum output in the saturation limited case is given by

$$I_{H\,\max}=\frac{E_{SH}/2}{R_O} \tag{66}$$

OUTPUT VERSUS SUPPLY VOLTAGE

The analysis used to determine the output versus supply voltage as shown in Fig. 7 is as follows:

m_1 = slope in the resistance-limited region of mode 1 operation obtained by differentiating equation 45 as

$$m_1=\frac{d(N_HI_H)}{dE_X}=\frac{N_X}{R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]+R_O\left(\frac{N_X}{N_H}\right)^2}\quad(\beta=\pi) \tag{67}$$

m_2 = slope in the region of mode 2 operation obtained by differentiating equation 59 as

$$m_2=\frac{d(N_HI_H)}{dE_X}=\frac{-N_X}{R_X+R_O\left(\frac{N_X}{N_H}\right)^2}\quad(\beta<\pi) \tag{68}$$

It is interesting to note that under these conditions of mode 2 operation the slope of the output-applied voltage curve is negative. The physical conditions leading to this phenomenon may be considered as follows. As the applied voltage is increased the saturable-reactor unit, in spite of an increased voltage, fires later in the cycle to maintain constant a-c ampere-turns for a constant signal. With an increasing voltage being applied to the transformer unit before firing, the transformer saturates earlier in the cycle. Since the load receives voltage only during the interval starting after the saturable-reactor unit fires and ending with the saturation of the transformer unit, it may be seen that under these conditions the output interval is shortened resulting in an output which decreases with increasing applied voltage.

The saturation limit to the output is obtained from the equation

$$N_H I_{H\,\max} = \frac{N_X E_{SX}}{2R_O\left(\frac{N_X}{N_H}\right)^2} \qquad (69)$$

The output during signal-limited operation is obtained from equation 44 as

$$N_H I_H = \frac{N_C I_C}{\left[2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]} \qquad (70)$$

The applied voltage at which the output becomes zero may be obtained by equating equation 59 to zero. Thus

$$E_{XO} = E_{SX} + \frac{N_C}{N_X} \times I_C R_X \quad (\chi > 1) \qquad (71)$$

The slope of the transition voltages locus as shown in Fig. 7 may be obtained by differentiating equation 61 as

$$\frac{dE_{X\ \mathrm{trans}}}{d(N_C I_C)} = \frac{-R_O\left(\frac{N_X}{N_H}\right)^2 + R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]}{N_X\left[2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]} \qquad (72)$$

Output Versus Frequency

Assuming that the core material has a unique saturating flux density independent of frequency, the effect of change of frequency may be viewed as a change of saturation voltage, E_{SX}. Then, for high frequencies the operation will be mode 1, either signal-limited or resistance-limited, and the output will be independent of frequency above the frequency where $\chi < 1/2$.

If, at high frequencies, the mode 1 operation is signal-limited the output will drop off at low frequencies when mode 2 operation is reached as shown by the dotted curve in Fig. 8 and as determined from equation 59 as follows:

$$\frac{d(N_H I_H)}{dE_{SX}} = \frac{N_X}{R_X + R_O\left(\frac{N_X}{N_H}\right)^2} = m_{\mathrm{sig}_2} \qquad (73)$$

If, on the other hand, the mode 1 operation is resistance-limited the output will drop off as mode 2 saturation-limited operation. The slope of the output versus supply voltage curve under these conditions is obtained by differentiating equation 69 as

$$\frac{d(N_H I_H)}{dE_{SX}} = \frac{N_X}{2R_O\left(\frac{N_X}{N_H}\right)^2} = m_{\mathrm{sat}} \qquad (74)$$

Although at intermediate frequencies the slope of the output-frequency curve will depend upon signal size, at very low frequencies the operation will become saturation-limited.

A consideration of equations 73 and 74 in the limit reveals that

$$m_{\mathrm{sig}_2} < 2m_{\mathrm{sat}_1} \qquad (75)$$

The point at which mode 2 operation changes from a signal-limited to a saturation-limited nature may be obtained by rearranging equation 61. Thus, by algebra,

$$E_{SX_{\mathrm{trans}}} = E_X - \left(\frac{N_C}{N_X}\right) I_C R_X \times \left[\frac{\left(\frac{R_O}{R_M}-\frac{R_O}{R_M}\right)\left(\frac{N_X}{N_H}\right)^2 - 1}{2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2}\right] \qquad (76)$$

From equation 61 it may be determined that mode 2 operation begins when

$$E_{SX_{\mathrm{trans}}}\ E_X + \frac{N_C I_C}{N_X} \times \frac{\left\{R_O\left(\frac{N_X}{N_H}\right)^2 - R_X\left[1+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]\right\}}{\left[2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]} \qquad (77)$$

The frequency below which the output is zero may be obtained by rearranging equation 71 as

$$E_{SX_0} < E_X - \frac{N_C}{N_X} I_C R_X \qquad (78)$$

Current Gain Versus Load Resistance

Underexcited Case

When R load is very large, α will equal 0 degrees, resistance-limited operation will ensue, and I_H will be governed by equation 45. On the other hand, when R load is not large the output is signal- and load-dependent as given by equation 44.

Overexcited Case

When R is large, α will equal 0 degrees and the saturation-limited output will be given by equation 68. However, when R is not large and operation is within the signal-controlled region of mode 2 operation, the output will be given by equation 59. From these four equations the current-gain versus load-resistance relation may be constructed as shown in Fig. 9 and as derived in the following.

In the region of high load resistance the curve of Fig. 9 may be obtained from the derivative of equation 45 or equation 69. Thus

$$m = \frac{d(N_H I_H)}{d(N_C I_C)} = 0 \qquad (79)$$

Similarly, in the region of mode 1 signal control from equation 44.

$$\frac{d(N_H I_H)}{d(N_C I_C)} = \frac{1}{\left[2+\frac{R_O}{R_M}\left(\frac{N_X}{N_H}\right)^2\right]} \qquad (80)$$

Similarly in the region of mode 2, in the saturation-limited region from the derivative of equation 69.

$$\frac{d(N_H I_H)}{d(N_C I_C)} = 0 \qquad (81)$$

In the signal-limited region, by differentiating equation 59,

$$\frac{d(N_H I_H)}{d(N_C I_C)} = \frac{1}{\left[1+\frac{R_O}{R_X}\left(\frac{N_X}{N_H}\right)^2\right]} \qquad (82)$$

Output Power

Defining the Thévenin equivalent impedance viewed from the output terminals as follows:

$$R_{TH} = R_H + \frac{R_M R_X\left(\frac{N_H}{N_X}\right)^2}{R_M + R_X} \qquad (83)$$

and defining the matching index n

$$n = \frac{R_{\mathrm{load}}}{R_{TH}} \qquad (84)$$

the maximum power output becomes

$$P_{L\max} = \frac{E_H{}^2 k_f{}^2}{4\left(1+\frac{R_X}{R_M}\right)^2 R_{TH}} \times \frac{4n}{(n+1)^2} \qquad (85)$$

The loading which results in maximum power transfer is obtained by differentiating the power equation and equating to zero. This results in $n=1$ for maximum power transfer.

References

1. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., first edition, 1955.

2. Series Connected Saturable Reactor with Control Source of Comparatively Low Impedance, H. F. Storm. *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 756–65.

3. Transient Response of Saturable Reactors with Resistive Load, H. F. Storm. *Ibid.* (*Electrical Engineering*), vol. 70, pt. I, 1951, pp. 95–102.

4. Magnetic-Amplifier Circuits (book), W. A. Geyger. McGraw-Hill Book Company, Inc., New York-Toronto-London, first edition, 1954.

5. Waveforms (book), B. Chance, V. Hughes, E. F. MacNichol, D. Sayre, F. C. Williams. McGraw-Hill Book Company, Inc., 1949.

6. Electronic Instruments (book), Ivan A. Greenwood, Jr., J. Vance Holdan, Jr., Duncan Macrae, Jr. McGraw-Hill Book Company, Inc., first edition, 1948.

7. Basic Facts About Magnetic Amplifiers, Harold A. Goldsmith, Herbert Hertz, Barney J. O'Neill. *Control Engineering*, New York, N. Y., vol. 1, no. 4, Dec. 1954, pp. 40–47.

8. Experimental Magnetic-type D-C to A-C Signal Inverter, J. Kauke. *Radio and Television News*, Chicago, Ill., vol. 43, May 1950, pp. 64–65.

9. DC/AC Converter, A Sensitive Method of Measuring Small DC Voltages, T. A. Ledward. *Wireless World*, London, England, vol. 48, Aug. 1943, pp. 230–33.

10. Magnetic Amplifiers for Airborne Applications, J. K. McKendry. *Tele-Tech*, Philadelphia, Pa., vol. XII, July 1953, pp. 32–34, 123, 125.

11. Use of Specially Designed Magnetic Amplifier in Computing-Circuits, H. S. Sack, R. T. Beyer, G. Miller, J. W. Trischka. *Report 437, Division 14*, National Defense Research Committee, Ithaca, N. Y., May 10, 1945.

12. The Magnetic Modulator in A-C Servo Corrective Networks, Carl Volz. *Proceedings*, The National Electronics Conference, Chicago, Ill., vol. IX, Sept. 1953, pp. 209–14.

13. A Simple Magnetic Modulator for Conversion of Millivolt D-C Signals, Gunnar Wennerberg. *Electrical Engineering*, vol. 70, Feb. 1951, pp. 144–47.

14. Regulating System for Alternating Current Circuits, **E. J. Murphy, L. W. Thompson.** *U. S. Patent No. 1,337,942*, Apr. 20, 1920.

15. Modulating System, **R. V. L. Hartley.** *U. S. Patent No. 1,462,038*, July 17, 1923.

16. Electrical Translating Apparatus, **A. J. Sorensen, et al.** *U. S. Patent No. 1,914,220*, June 13, 1933.

17. Regulating System, **O. Schmutz.** *U. S. Patent No. 1,997,657*, Apr. 16, 1935.

18. Magnetic Modulator, **R. Lee.** *U. S. Patent No. 2,265,296*, Dec. 2, 1947.

19. Automatic Control System, **P. A. Noxon.** *U. S. Patent No. 2,432,036*, Dec. 2, 1947.

20. Control Circuit for Induction Motors, **W. J. Krupick.** *U. S. Patent No. 2,435,926*, Feb. 10, 1948.

21. Electric Motor Servo Control System, **J. F. Emerson.** *U. S. Patent No. 2,450,084*, Sept. 28, 1948.

22. Control System, **H. Conviser.** *U. S. Patent No. 2,487,697*, Nov. 8, 1949.

23. Electric Motor Control System, **H. Conviser.** *U. S. Patent No. 2,545,343*, Mar. 13, 1951.

24. Servo System with Saturable Core Component, **J. P. Palmer.** *U. S. Patent No. 2,559,513*, July 3, 1951.

25. Induction Motor Control System, **A. M. MacCullum, et al.** *U. S. Patent No. 2,608,676*, Aug. 26, 1952.

26. Harmonic Discriminating Cathode Follower Circuit in a Servo System, **O. N. Perkins, Sr., et al.** *U. S. Patent No. 2,676,291*, Apr. 20, 1954.

27. Saturable Transformer Device, **A. Bennett.** *U. S. Patent No. 2,678,419*, May 11, 1954.

28. Saturable Transformer Amplifier, **G. B. Houck, Jr.** *U. S. Patent No. 2,679,621*, May 25, 1954.

◆

Discussion

E. W. Lehtonen (Sperry Gyroscope Company, New Hyde Park, N. Y.): The circuit in Fig. 2 of the paper consists of two magnetic structures, and if of ideal core material and operated in mode 1, signal-limited, they must obey the equal ampere-turn law. I am assuming that both control windings are carrying direct current, the current in the upper half being I_{ca} and the other I_{cb}.

$$I_xN_x-I_hN_h=I_{ca}N_c \tag{89}$$

$$I_xN_x+I_hN_h=I_{cb}N_c \tag{90}$$

Subtract equation 89 from equation 90 and divide by 2

$$I_hN_h=N_c(I_{cb}-I_{ca})/2 \tag{91}$$

This equation corresponds to equation 44 of Mr. Anderson's paper except:

1. Equation 44 assumes $I_{ca}=0$.

2. Equation 91 above lacks the term $(R_o/R_m)(N_x/N_h)^2$, since ideal core material was assumed. This term is necessary only for the sake of completeness, but usually is quite small. Empirically, I have found that changing from 50 per-cent nickel-steel to silicon steel reduced gain by only 5 per cent. The degree of difference of course depends on how large the d-c magnetomotive force is, compared to the coercive force of the core, so that the size of the unit will affect the accuracy of the assumption that the core material is ideal. Add equations 89 and 90 and divide by 2.

$$I_xN_x=N_c(I_{ca}+I_{cb})/2 \tag{92}$$

Equation 92 defines the current drawn from the exciting source. Resistance- and signal-limited conditions are naturally the same as shown in the paper, except for the slight difference caused by neglect of core loss.

A study of equation 91 leads to the conclusion that the circuit behaves like a pair of saturable reactors in push-pull as shown in Fig. 19. The symbols correspond to those of Mr. Anderson's paper as far as possible. The resistances $(R_x(N_h/N_x)^2+R_h)$ of the saturable reactors are twice as large as would be realized if the reactors were wound for the same current density on the same cores as the equivalent saturable transformers. That this circuit is equivalent to push-pull saturable transformers was proved in the laboratory by constructing a unit with $N_x=N_h$ so that the same coils could be used in either circuit. The relationship between load current and signal currents was identical in both cases. The purpose of mentioning this equivalent circuit is that it makes more obvious the fact that the output impedance is very high. The output impedance given in equation 83 only applies when the load current is resistance-limited. Whenever the load current is determined by signal ampere-turns (e.g., equation 91, or equation 44 of the paper), the load sees a current source and therefore a high impedance. This high impedance is important in that a servo motor would not experience much damping over most of the control range. A further consideration is that matching the load in accordance with equation 83 will give rise to serious thermal problems in all but the smallest units if the signal currents were increased to the point where the load current is resistance-limited as assumed by equation 83.

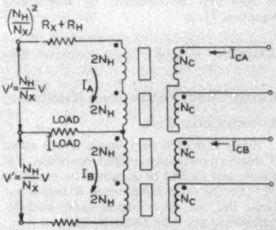


Fig. 19. Saturable-reactor circuit equivalent to push-pull saturable transformer

The equivalent circuit also points up that the restriction of signal circuit resistances to low values is unnecessary as far as average ampere-turns are concerned. The wave forms would of course be affected. I would like to know if Mr. Anderson has found any practical use for mode 2 operation.

I wish Mr. Anderson well with the problems of leakage inductance and reactive loads. I particularly hope for a criterion for instability with capacitive loads. A capacitive load with leakage inductance can lead to an increase of gain, sometimes enough to result in instability. This type of behavior has been discussed by J. M. Manley.[1]

I would like also to point out that in Fig. 17, the circuit is called a "magnetic modulator." While this circuit performs as a modulator, the term "magnetic modulator" is usually restricted to the small signal case where the equal ampere-turn law is not applicable, and the wave forms are quite different.

Reference

1. Some General Properties of Magnetic Amplifiers, **J. M. Manley.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., Mar. 1951.

R. E. Anderson: Mr. Lehtonen's interesting and stimulating discussion is certainly appreciated. As he indicates, the difference between his alternate solution and that of the paper is concerned with the width of the assumed *B-H* loop. In the alternate solution of the discussion the loop width is considered negligible or zero, while the paper included a core loss, Fig. 3.

Accordingly it may be of interest to mention the reasons for introducing the magnetizing term, R_M, into the analysis. Certain manufacturing experiences indicated that parameters other than turns and resistances influenced performance. The core-loss term which I used, R_M, enabled performance prediction with sufficient accuracy to establish the desired quality control. For instance, in this circuit the gain is a function of loading, a consideration most important in design. This loading dependency, explained in Fig. 9, is a manifestation of the core loss. In the example of Figure 11(A), approximately 50-per-cent gain improvement may be obtained with 50-per-cent nickel-iron material. Nevertheless, it should be clearly pointed out that many other considerations, such as saturation flux density, may properly determine the choice of core material in a practical design. The experimental transformer described in the paper was intended only as a means for verifying the analysis.

I am glad to see Mr. Lehtonen express an interest in mode 2 operation because this circuit characteristic, shown in Fig. 7, is seldom encountered in other devices. One of the obvious applications would be the construction of a static switching device for both over- and undervoltage protection. Actually, circuits have been constructed where mode 2 operation is obtained for small signals with transition to mode 1 taking place at larger signal levels. Interestingly enough, most saturable-core transformers available commercially today are designed to reduce size and weight by operating near normal excitation at nominal conditions of voltage and frequency. Then, of course, such designs are operating in mode 2 at the extremes of voltage and frequency encountered in typical military applications.

It is heartening to have Mr. Lehtonen emphasize the leakage reactance problem as an important area for future work. Perhaps this will encourage contributions in this area. In regard to the instability problem arising with capacitive loading, one of my colleagues, Glenn Walters, calls attention to the possibility of using a technique similar to the method of Mandelstam, described by Minorsky,[1] for analyzing this condition. I would like to thank Mr. Lehtonen for the important addition to the bibliography, for the numerous items of information he made available, for the equivalent circuit he presents, and for the stimulating discussions which he prepared.

REFERENCE

1. INTRODUCTION TO NON-LINEAR MECHANICS (book), N. Minorsky. J. W. Edwards Company, Ann Arbor, Mich., 1947.

Upper Limits of Output Power in Vacuum-Tube and Transistor A-C Amplifiers

L. M. VALLESE
ASSOCIATE MEMBER AIEE

Synopsis: It is shown that the upper limits of output power of a-c vacuum-tube and transistor amplifiers are determined by certain characteristic device values which depend on their maximum voltage, current, and power dissipation.

THE PURPOSE of this paper is to investigate the variation of the upper limits of the output power of a-c amplifiers as a function of the amplitude of the a-c input signal or of that of the d-c output voltage. These upper limits depend on the current-voltage characteristics of the amplifying device and on their boundaries of maximum voltage, current, and power dissipation. Assuming idealized characteristics, simple analytical relationships may be derived which illustrate such a dependence. These also provide a quick design of first approximation; indicate the relative importance of the voltage, current, and power limitations; and afford a comparison of performance for different amplifiers.

Single-ended (class *A*) and push-pull (class *B*) transformer-coupled configurations are considered.

Vacuum-Tube Triodes

The output characteristics are idealized as shown in Fig. 1:

$$I_b = g_m v_{ck} + g_p v_{bk} \qquad (1)$$

$$0 \geq v_{ck}$$

$$0 \leq v_{bk} \leq V_{b\,\max}$$

$$0 \leq I_b \leq I_{b\,\max}$$

The condition of maximum permissible dissipation is the following equation

$$V_{bQ} I_{bQ} \leq P_p + P_L \qquad (2)$$

where V_{bQ}, I_{bQ} are quiescent values, P_p and P_L are respectively the maximum permissible plate dissipation and the output power. For class *A* amplifiers the relation in equation 2 reduces to

$$V_{bQ} I_{bQ} \leq P_p \qquad (3)$$

In general, vacuum-tube manufacturers give the value of $V_{bQ\,\max}$ rather than that of $V_{b\,\max}$.

Consider a single-ended transformer-coupled amplifier. With the dynamic load line terminated respectively at the $v_{ck}=0$ and $I_b=0$ lines, the following equations apply:

$$I'' = 2I_{bQ} = g_p(E_{bb} - E_m) \qquad (4)$$

$$I_{bQ} = G_L' E_m = -g_m E_{gm} + g_p E_{bb}$$

where I'' is the peak I_b value, E_{gm} and E_m are the amplitudes of the input and output alternating voltages and G_L' the primary transformed load admittance. From equations 1 and 4 there follows

$$E_m = \frac{g_m E_{gm}}{g_p + G_L'} = \frac{g_p E_{bb}}{g_p + 2G_L'} \qquad (5)$$

When E_{gm} is prescribed and is small enough that the $I_{b\,\max}$ and the P_p limitations are not involved, maximization of P_L provides the well-known results

$$G_L' = g_p,\ E_{bb} = \frac{3\mu}{2} E_{gm},\ E_m = \frac{\mu}{2} E_{gm} \qquad (6)$$

$$I_{bQ} = \frac{g_m}{2} E_{gm},\ P_L = \frac{\mu g_m}{8} E^2_{gm},\ \eta = 0.166$$

However, if

$$E_{gm} \geqq E_{gm}{}^* = \frac{I_{b\,\max}}{g_m} \text{ or}$$

$$E_{gm} \geqq E_{gm}{}^{**} = \frac{2}{g_m}\sqrt{\frac{g_p P_p}{3}} \qquad (7)$$

in which case the $I_{b\,\max}$ or the P_p limitations apply respectively, equations 6 do not apply. Assuming that $E_{gm}{}^* < E_{gm}{}^{**}$, a condition often verified in practice, the dynamic load line must be terminated at point *A* of Fig. 1 for $E_{gm} \geqq E_{gm}{}^*$. Correspondingly, it is found

$$G_L' = g_p/(2E_{gm}/E_{gm}{}^* - 1),$$

$$E_{bb} = \frac{I_{bM}}{2g_p}\left(1 + \frac{2E_{gm}}{E_{gm}{}^*}\right) \qquad (8)$$

$$E_m = E_{bb} - 2P_p/g_p E_{bb},\ I_{bQ} = P_p/E_{bb}$$

$$P_L = \frac{I_{bm}{}^2}{8g_p}(2E_{gm}/E_{gm}{}^* - 1),$$

$$\eta = \frac{1}{2}\,\frac{2E_{gm}/E_{gm}{}^* - 1}{2E_{gm}/E_{gm}{}^* + 1}$$

When

$$E_{gm} \geqq E_{gm}{}^{***} = \left(\frac{2g_p}{I_{bm}{}^2} P_p - 1/2\right) E_{gm}{}^* \qquad (9)$$

the P_p limitation occurs and the following design equations apply instead of equations 8:

$$G_L' = g_p \frac{P_p}{g_p E_{bb}{}^2 - 2P_p},\ E_{bb} = \mu E_{gm}/2 + \sqrt{(\mu E_{gm}/2)^2 + P_p/g_p}$$

$$E_m = E_{bb} - 2P_p/g_p E_{bb},\ I_{bQ} = P_p/E_{bb} \qquad (10)$$

$$P_L = \frac{P_p}{2}(1 - 2P_p/g_p E_{bb}{}^2),$$

$$\eta = \frac{1}{2}(1 - 2P_p/g_p E_{bb}{}^2)$$

Note that, if $E_{gm}{}^* > E_{gm}{}^{**}$, the results of equation 10 apply for $E_{gm} \geqq E_{gm}{}^{**}$. The maximum permissible E_{gm} value is $E_{gm\,\max} = (1/\mu)(E_{bb\,\max} - P_p/g_p E_{bb\,\max})$.

Paper 57-74, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted September 24, 1956; made available for printing December 6, 1956.

L. M. VALLESE is with the Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

This work was done at the Microwave Research Institute of the Polytechnic Institute of Brooklyn, under a contract sponsored by the Office of Naval Research.

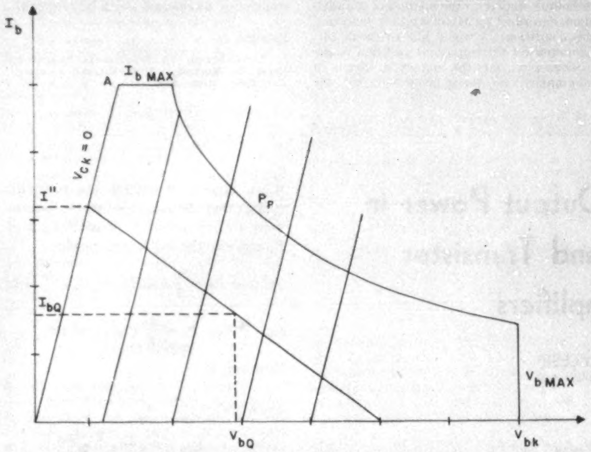


Fig. 1. Idealized characteristics of a vacuum-tube triode

An illustrative example is shown in Fig. 2 for a *6AS7G* triode for which $I_{b\,\max}=125$ ma, $V_{bQ}=250$ v(volts), $P_p=13$ watts, $g_p=3.57\times10^{-3}$ mhos, $g_m=7\times10^{-3}$ mhos. In this case it is found that $E_{gm}{}^*=17.9$ v, $E_{gm}{}^{**}=35.4$ v, $E_{gm}{}^{***}=97.5$ v, $E_{gm\,\max}=117.7$ v.

Consider now the case in which E_{bb} is given. If the $I_{b\,\max}$ and P_p limitations are not involved, maximization of P_L provides the well-known results

$$G_L'=g_p/2,\ E_{gm}=3E_{bb}/4\mu,\ E_m=E_{bb}/2 \qquad (11)$$
$$I_{ba}=g_pE_{bb}/4,\ P_L=g_pE_{bb}{}^2/16,\ \eta=0.25$$

Their validity ceases when

$$E_{bb}\geqq E_{bb}{}^*=2I_{bM}/g_p \text{ or } E_{bb}\geqq E_{bb}{}^{**}=2\sqrt{P_p/g_p} \qquad (12)$$

for which cases the I_{bM} or the P_p limitations occur respectively. If $E_{bb}{}^*<E_{bb}{}^{**}$, the dynamic load line is terminated at point *A* of Fig. 1 for $E_{bb}\geqq E_{bb}{}^*$ and correspondingly it is found that

$$G_L'=\frac{g_p}{2}\Big/\left(2\frac{E_{bb}}{E_{bb}{}^*}-1\right),\ E_{gm}=\frac{1}{\mu}\times\left(E_{bb}-\frac{E_{bb}{}^*}{4}\right),\ E_m=E_{bb}-\frac{E_{bb}{}^*}{2} \qquad (13)$$

$$I_{bQ}=I_{bm}/2,\ P_L=g_p\frac{E_{bb}{}^*}{8}\left(E_{bb}-\frac{E_{bb}{}^*}{2}\right),\ \eta=\frac{1}{2}\left(1-\frac{E_{bb}{}^*}{2E_{bb}}\right)$$

Finally, when

$$E_{bb}\geqq E_{bb}{}^{***}=2P_p/I_{bM} \qquad (14)$$

the P_p limitation occurs and the design equations are

$$G_L'=\frac{g_p}{2}\Big/\left(g_p\frac{E_{bb}{}^2}{2P_p}-1\right),\ E_{gm}=\left(E_{bb}-\frac{P_p}{g_pE_{bb}}\right)\Big/\mu \qquad (15)$$

The expressions of E_m, I_{bQ}, P_L, and η are the same as in equation 10.

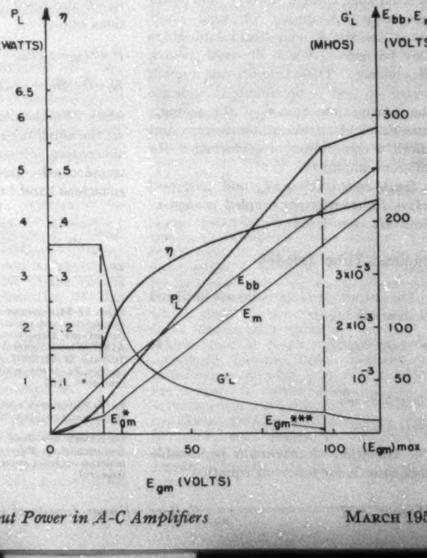


Fig. 2. Idealized optimum design data for 6ASG7 amplifier with prescribed E_{gm}

An illustrative design for a *6AS7G* triode is shown in Fig. 3, where $E_{bb}{}^*=70$ v, $E_{bb}{}^{**}=120.4$ v, $E_{bb}{}^{***}=208$ v, $E_{bb\,\max}=250$ v.

The design of a class *B* push-pull amplifier is obtained on the basis of the equations

$$I_b=g_mv_{gk}+g_p(v_{bk}-E_{bb})$$
$$I_m=G_L'E_m=g_mE_{gm}-g_pE_m \qquad (16)$$
$$I_m\leqq I_{b\,\max}$$
$$E_{bb}I_{bQ}-E_mI_m/4\leqq P_p$$

where v_{gk} is the a-c input voltage, G_L' the plate-to-cathode load admittance, I_m the peak a-c output current, and P_p the plate dissipation of each tube.

If the I_{bM} and the P_p limitations are not involved, maximizing P_L (power output for two tubes) gives the following known results:

$$G_L'=g_p,\ E_{gm}=E_{bb}/\mu,\ E_m=E_{bb}/2 \qquad (17)$$
$$I_{bQ}=I_m/\pi=g_pE_{bb}/2\pi,\ P_L/2=g_pE_{bb}{}^2/16,\ \eta=\pi/8$$

On the other hand, if

$$E_{bb}\geqq E_{bb}{}^*=2I_{bM}/g_p,\text{ or } E_{bb}\geqq E_{bb}{}^{**}=3.2\sqrt{P_p/g_p} \qquad (18)$$

in which case the I_{bM} or the P_p limitations respectively apply, the design equations must be modified. In general, $E_{bb}{}^*<E_{bb}{}^{**}$; accordingly for $E_{bb}\geqq E_{bb}{}^*$, letting $I_m=I_{bM}$, the following design equations are found:

$$G_L' = g_p\Big/\left(2\frac{E_{bb}}{E_{bb}^*}-1\right),\ E_{gm}=E_{bb}/\mu,$$
$$E_m = E_{bb} - E_{bb}^*/2 \tag{19}$$

$$I_{bQ}=I_{bM}/\pi,\ P_L/2=I_{bM}\times \left(E_{bb}-\frac{E_{bb}^*}{2}\right)\Big/4,\ \eta=\frac{\pi}{4}\left(1-\frac{E_{bb}^*}{2E_{bb}}\right)$$

If

$$E_{bb} \geqq E_{bb}^{***} = 14.6\left(\frac{P_p}{I_{bM}}-\frac{E_{bb}^*}{8}\right) \tag{20}$$

or if $E_{bb}^* > E_{bb}^{**}$ and $E_{bb} \geqq E_{bb}^{**}$, the following design equations apply:

$$G_L' = g_pI_m/(g_pE_{bb}-I_m),\ I_m = 29.2g_pE_{bb}+\sqrt{(29.2g_pE_{bb})^2+4g_pP_p}$$
$$E_{gm}=E_{bb}/\mu,\ E_m=E_{bb}-I_m/g_p,\ I_{bQ}=I_m/\pi \tag{21}$$

$$P_L/2=I_m\left(E_{bb}-\frac{I_m}{g_p}\right)\Big/4,\ \eta=P_L/2P_p$$

In general, $E_{bb}^{***} >> E_{bb\ \max}$.

An illustrative example is shown in Fig. 4 for two *6AS7G* triodes in class *B* push-pull configuration. In this case $E_{bb}^* = 70$ v, $E_{bb}^{**} = 193$ v, $E_{bb}^{***} = 1{,}395$ v.

In conclusion, the upper limit of the power output of a single-ended transformer coupled triode amplifier increases monotonically with E_{gm} or E_{bb}; its peak value is $P_p(1-2P_p/g_pE_{bb\ \max}^2)/2$. The upper limit of power output for a push-pull class *B* amplifier also increases monotonically with E_{bb}; its peak value (for two tubes) is generally $I_{bM}(E_{bb\ \max}-I_{bM}/g_p)/2$. If the design is limited by P_p rather than I_{bM}, the peak value of P_L follows from equations 21 on letting $E_{bb}=E_{bb\ \max}$. Such value tends asymptotically to 7.3 P_p.

Pentodes and Beam Power Tetrodes

The idealized characteristics of pentodes and beam power tetrodes considered in this study are shown in Fig. 5. Their analytical representation is

$$I_b = I_0 + g_mv_{ck}$$
$$0 \geqq v_{ck} \tag{22}$$
$$v_b^0 \leqq v_{bk} \leqq v_{b\ \max}$$
$$0 \leqq I_b \leqq I_0$$

In addition, equations 2 and 3 also apply.

The design of single-ended transformer-coupled amplifiers for maximum power output is obtained from the equations

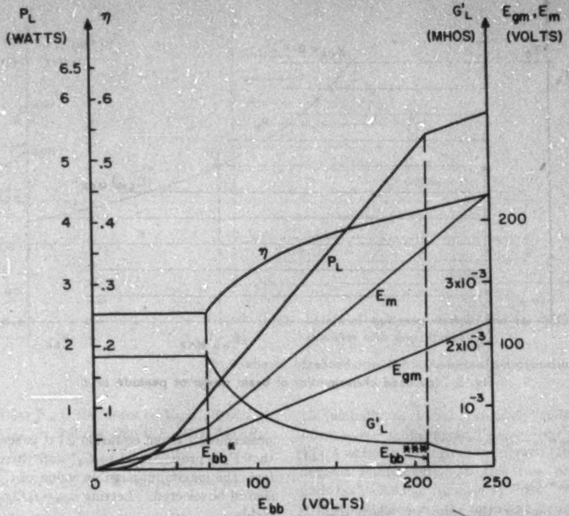


Fig. 3. Idealized optimum design data for 6ASG7 amplifier with prescribed E_{bb}

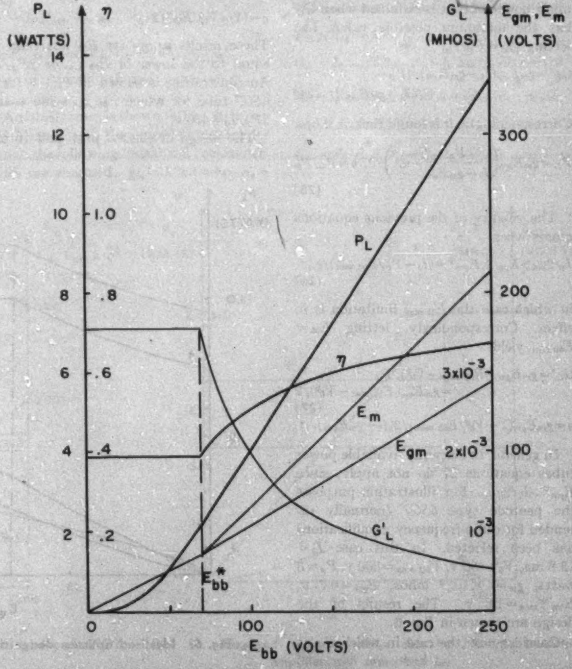


Fig. 4 (right). Idealized optimum design data for class B push-pull 6ASG7 amplifier

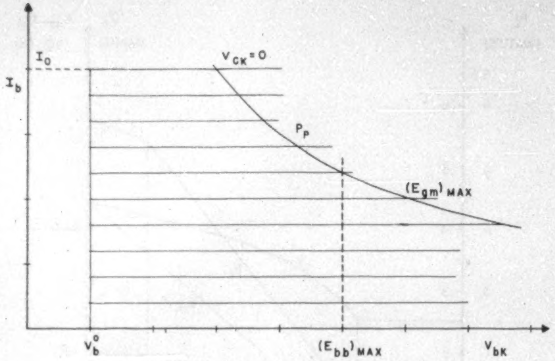


Fig. 5. Idealized characteristics of beam power or pentode tube

$$I_{bQ}=I_0-G_L'E_m=I_0-g_mE_{gm}$$
$$E_m=E_{bb}-v_b^0 \qquad (23)$$

Consider the case in which E_{gm} is prescribed, with $0\leqq E_{gm}\leqq E_{gm\ \max}=I_0/2g_m$. Since P_L varies inversely with G_L', maximum power output is obtained when G_L' has the minimum possible value, i.e., letting $E_{bb}=P_p/I_{bQ}$

$$G_L'=g_mE_{gm}(I_0-g_mE_{gm})/[P_p-V_b^0(I_0-g_mE_{gm})] \qquad (24)$$

Correspondingly, it is found that

$$P_L=g_mE_{gm}\left(\frac{P_p}{I_0-g_mE_{gm}}-v_b^0\right)/2,\ \eta=P_L/P_p \qquad (25)$$

The validity of the previous equations ceases when

$$I_0/2g_m>E_{gm}\geqq E_{gm}^*=(I_0-P_p/E_{bb\ \max})/g_m \qquad (26)$$

in which case the $E_{bb\ \max}$ limitation is in effect. Correspondingly, letting $E_{bb}=E_{bb\ \max}$, yields

$$G_L'=g_mE_{gm}/(E_{bb\ \max}-V_b^0),\ P_L=g_mE_{gm}(E_{bb\ \max}-V_b^0)/2$$
$$\eta=g_mE_{gm}(1-V_b^0/E_{bb\ \max})/2(I_0-g_mE_{gm}) \qquad (27)$$

In general, in currently available power tubes equations 27 do not apply, since $E_{gm}^*>I_0/2g_m$. For illustration purposes the pentode type *6SG7* (normally intended for radio-frequency amplification) has been selected. In this case $I_0=12.8$ ma, $V_b^0=50$ v, $V_{bQ\ \max}=300$ v, $P_p=3$ watts, $g_m=4\times10^{-3}$ mhos, $E_{gm}=0.7$ v, $E_{gm\ \max}=1.6$ v. The results of the design are shown in Fig. 6.

Consider now the case in which E_{bb} is prescribed. From equation 23 it is seen that P_L is proportional to G_L' and therefore the maximum possible value of G_L' should be selected. Letting $E_{gm}=I_0/2g_m$, yields

$$G_L'=I_0/2(E_{bb}-V_b^0),\ P_L=I_0(E_{bb}-V_b^0)/4$$
$$\eta=(1-V_b^0/E_{bb})/2 \qquad (28)$$

These results apply for E_{bb} less than or equal to the lower of $E_{bb\ \max}$ or $2P_p/I_0$. An illustration is shown in Fig. 7 for a *6SG7* tube for which $E_{bb\ \max}=300$ v and $2P_p/I_0=470$ v.

The design of class *B* push-pull amplifiers is obtained with the two equations

$$I_b=g_mv_{gk},\ I_m=I_0$$
$$E_m=E_{bb}-V_b^0 \qquad (29)$$

There follows

$$E_{gm}=I_0/g_m,\ G_L'=I_0/(E_{bb}-V_b^0)$$
$$P_L=I_0(E_{bb}-V_b^0)/2,\ \eta=\frac{\pi}{4}(1-V_b^0/E_{bb}) \qquad (30)$$

where P_L is shown for two tubes. The upper limit of E_{bb} is the lower of $E_{bb\ \max}$ or $14.6\ (P_p/I_0+V_b^0/4)$; usually the latter quantity is much larger than $E_{bb\ \max}$.

For example, in Fig. 8 is summarized the design of a push-pull class *B* amplifier with *6SG7* pentodes, for which $14.6\ (P_p/I_0+V_b^0/4)=3{,}600$ *V*.

In conclusion, the upper limit of the output power of a pentode single-ended or push-pull amplifier increases monotonically with E_{gm} or E_{bb}. For a single-ended amplifier the peak power is either $P_p/2-V_b^0I_0/4$ or $I_0(E_{bb\ \max}-V_b^0)/4$ accordingly, as equations 25 or 27 apply. For a class *B* push-pull amplifier the peak power output (two tubes) is $I_0(E_{bb\ \max}-V_b^0)/2$.

Transistors

Assuming for convenience of reference a p-n-p type of transistor in the common emitter configuration, its idealized characteristics may be written (Fig. 9)

$$I_b=g_{bb}V_{be},\ I_c=\alpha_eI_b=g_mV_{be}$$

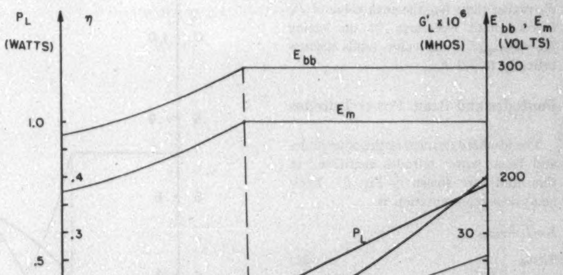
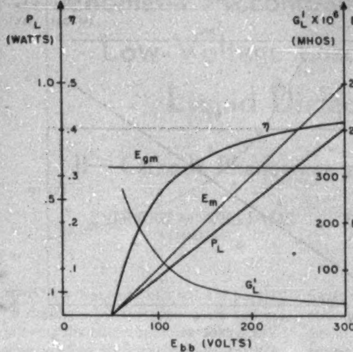


Fig. 6. Idealized optimum design data for 6SG7 amplifier with prescribed E_{gm}



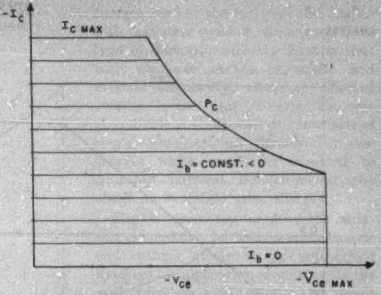


Fig. 7 (left). Idealized optimum design data for 6SG7 amplifier with prescribed E_{bb}

Fig. 9 (above). Idealized output characteristics of p-n-p transistor

$$0 \geqq I_c \geqq I_{c\,\max},\ 0 \geqq V_{ce} \geqq V_{ce\,\max} \tag{31}$$

In addition, $V_{eQ}I_{eQ} \leqq P_c + P_L$ and $V_{eQ}I_{eQ} \leqq P_c$ replace respectively equations 2 and 3.

The maximum power output design of a single-ended transformer-coupled amplifier is obtained from the equations

$$\begin{aligned} I_c'' &= 2\alpha_e I_{bn} = 2I_{cQ} \\ I_{cQ} &= G_L'E_m = G_L'V_{cQ} = \alpha_e I_{bm} \end{aligned} \tag{32}$$

where I_{bm} and E_m are respectively the amplitudes of the a-c input signal and a-c output voltage.

For prescribed I_{bm}, maximum power output is obtained when G_L' is as small as possible. Since

$$G_L' = \alpha_e I_{bm}/E_{cc} \tag{33}$$

where E_{cc} is the lower of $V_{ce\,\max}/2$ or $P_c/\alpha_e I_{bm}$, it follows that if $I_{bm} \leqq I_{bm}^* = 2P_c/\alpha_e V_{ce\,\max}$

$$\begin{aligned} &G_L' = \alpha_e^2 I_{bm}^2/P_c,\ P_L = \alpha_e I_{bm} V_{ce\,\max}/4, \eta = 1/2 \\ &E_{cc} = V_{ce\,\max}/2,\ P_L/P_{in} = g_m V_{ce\,\max}/2I_{bm} \end{aligned} \tag{34}$$

and if $I_{bm} \geqq I_{bm}^*$

$$\begin{aligned} &G_L' = 2\alpha_e I_{bm}/V_{ce\,\max},\ P_L = P_c/2,\ \eta = 1/2 \\ &E_{cc} = P_c/\alpha_e I_{bm},\ P_L/P_{in} = g_m P_c/\alpha_e I_{bm}^2 \end{aligned} \tag{35}$$

The maximum possible value of I_{bm} is $I_{c\,\max}/2\alpha_e$.

An illustration is shown in Fig. 10 for an *H3* Minneapolis-Honeywell transistor for which the following "idealized" characteristics are assumed: $g_m = 0.055$ mho, $\alpha_e = 15$ (actually α_e varies "inversely" with I_c), $I_{c\,\max} = -0.35$ ampere, $V_{ce\max} = -60$ v, $P_c = 4$ watts at 40 degrees centigrade (100 degrees Fahrenheit). In this case one finds $I_{bm}^* = 8.9$ ma, $I_{bm\,\max} = 11.7$ ma.

For prescribed E_{cc} the optimum design is obtained with G_L' as large as possible. From equation 33 where I_{bm} is the lower of $I_{c\,\max}/2\alpha_e$ or $P_c/\alpha_e E_{cc}$, it follows that if

$$E_{cc} \leqq E_{cc}^* = 2P_c/I_{c\,\max}$$

$$\begin{aligned} &G_L' = I_{c\,\max}/2E_{cc},\ P_L = I_{c\,\max}E_{cc}/4,\ \eta = 1/2 \\ &I_{bm} = I_{c\,\max}/2\alpha_e,\ P_L/P_{in} = 2g_m\alpha_e E_{cc}/I_{c\,\max} \end{aligned} \tag{36}$$

and if $E_{cc} \geqq E_{cc}^*$

$$G_L' = P_c/E_{cc}^2,\ P_L = P_c/2,\ \eta = 1/2$$

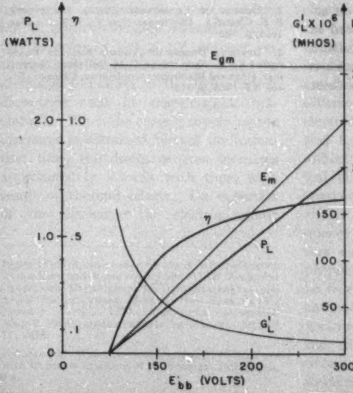


Fig. 8. Idealized optimum design data for class B push-pull 6SG7 amplifier

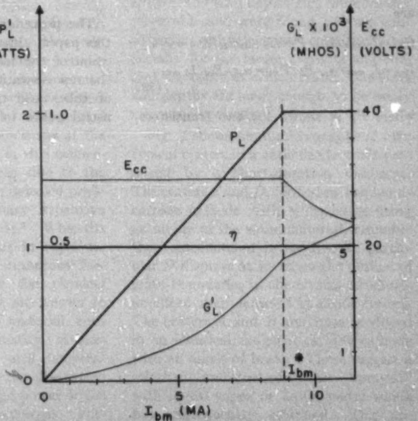


Fig. 10. Idealized optimum design data for H3-type transistor amplifier with prescribed I_{bm}

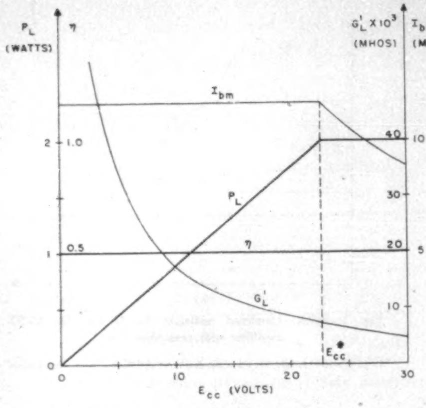


Fig. 11. Idealized optimum design data for H3 type transistor amplifier with prescribed E_{cc}

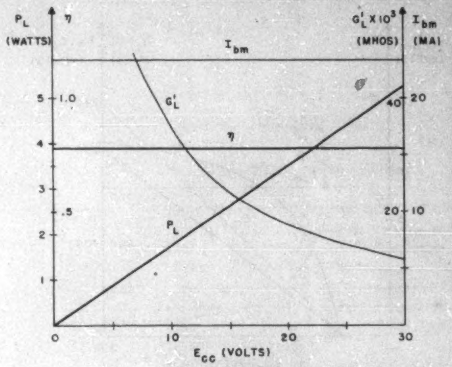


Fig. 12. Idealized optimum design data for push-pull, class B, H3-type transistor amplifier

$$I_{bm}=P_c/\alpha_e E_{cc},\; P_L/P_{in}=\alpha_e g_m E_{cc}^2/P_c \quad (37)$$

An illustration of this design is shown in Fig. 11 for the transistor type *H3*. In this case $E_{cc}^*=22.8$ v.

Finally, the design of a class *B* push-pull amplifier for maximum power output is obtained from the equations

$$I_c=\alpha_e i_b,\; I_{cQ}=\alpha_e I_{bm}/\pi,\; \alpha_e I_{bm}=G_L' E_m$$

where i_b is the a-c component of the input signal current. In this case the value of I_{bm} is the lower of $I_{c\max}/\alpha_e$ or 14.6 $P_c/\alpha_e E_{cc}$. There follows that, if $E_{cc} \leqq E_{cc}^{**}=14.6\, P_c/I_{c\,\max}$, the design equations are

$$G_L'=I_{c\,\max}/E_{cc},\; P_L=I_{c\,\max}E_{cc}/2,\; \eta=\pi/4$$

$$I_{bm}=I_{c\,\max}/\alpha_e,\; P_L/P_{in}=\alpha_e g_m E_{cc}/I_{c\,\max} \quad (38)$$

where P_L is shown for two transistors. If $E_{cc\,\max} \geqq E_{cc} \geqq E_{cc}^{**}$, the design equations are

$$G_L'=14.6\, P_c/E_{cc}^2,\; P_L=7.3\, P_c,\; \eta=\pi/4 \quad (39)$$

$$I_{bm}=14.6\, P_c/\alpha_e E_{cc},\; P_L/P_{in}= \alpha_e g_m E_{cc}^2/14.6\, P_c$$

An illustration is shown in Fig. 12 for two *H3* transistors, for which $E_{cc}=166\, V >> E_{cc\,\max}$. It appears that in this case the asymptotic value of $P_L=7.3\, P_c$ cannot be realized.

Conclusions

The procedure of design discussed in this paper, although approximate on account of the idealizations introduced in the representation of the characteristics of tubes and transistors, constitutes a useful guide to the choice of the best operating conditions. It appears that transistors perform better than tubes not only on account of their greater efficiency, but also because generally it is possible to realize with them the theoretical asymptotic value of output power.

In addition, the results of the analysis may be of interest for device engineers; as a matter of fact, they may be used as a guide for more efficient design of power tubes and transistors, as well as for a rational classification of types.

References

1. On the Optimum Design of Cathode Followers, L. M. Vallese. *AIEE Transactions*, vol. 74, pt. I, Mar. 1955, pp. 5–8.

2. Design of Transistor Power Amplifiers, S. K. Ghandhi. *Electronics*, New York, N. Y., Mar. 1954, p. 146.

3. Optimum Design of Common Emitter Transistor Audio Amplifiers, L. M. Vallese. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. XI, 1955, p. 919.

Phenomena Accompanying Transient Low-Voltage Discharges in Liquid Dielectrics

II—Cathode Phenomena at Low Currents

EVERARD M. WILLIAMS
MEMBER AIEE

RICHARD E. SMITH
ASSOCIATE MEMBER AIEE

THIS is the second of a series of papers[1] dealing with the phenomena accompanying discharges resulting from sparkover at low voltages between metallic electrodes in a liquid dielectric. These phenomena are of particular concern because of developments in the art of electric-spark machining[2] in which the effects of such discharges are used in the shaping of materials.

Sparkover of a short liquid dielectric path separating metallic electrodes, to which a suitable voltage is applied, takes place after a timelag which depends on statistical considerations as well as upon the voltage gradients involved. After the statistical time delay the actual formation time of the discharge is in the range of 10 to 20 millimicroseconds. When completely formed, the discharge path comprises an approximately cylindrical column in which substantially all the liquid material originally located in the path has been converted to high-temperature vapor under extremely high pressure. This pressure is maintained by the electromagnetic forces accompanying an abnormally high current density, (of the order of 100×10^6 amperes per square inch immediately following the discharge formation in the hydrocarbon dielectrics used in electric-spark machines). When the current supplying the discharge is sustained beyond the formation time, the discharge area increases approximately linearly with time, as a result of thermal effects. On cessation of the discharge the electromagnetic forces vanish and the high-pressure vapor column expands explosively. Throughout or at some time during the process, craters are formed in the anode and cathode electrodes. The material removed from the anode and cathode corresponds to the useful machining and incidental tool wear respectively in the electric-spark machine tool. The first paper[1] in this series treated the physical phenomena relating to anode-crater formation. This paper is concerned with physical phenomena at the cathode. A subsequent paper will deal with the discharge dynamics.

Previous investigators have studied electric-spark machining phenomena in the electric-spark machine tool itself, in which the cumulative effects of many discharges in rapid sequence are observed. One of the earliest results was the disclosure of a distinct difference in volume of material removed from anode and cathode, which led to the choice of tool (cathode) to workpiece (anode) polarity relation universally employed in electric-spark machine tools. However, the data in general have been very crude. It has not, for instance, been known whether the simple crater-volume differentials are accompanied by well-defined differences in physical phenomena at the electrodes. Lack of data in this connection has undoubtedly been due to the difficulty of distinguishing detailed physical effects in the ordinary repetitive sparkover-machining process.[3] When the effects of many discharges in rapid sequence are superposed, numerous factors combine to obscure the physical phenomena; among these are known to be redeposition of eroded material, both cathode and anode, formation of low-pressure vapor pockets, and temperature effects due either to the dissipation of large amounts of average power or the formation of continuous discharges. Furthermore, although emphasis has been placed on the distinction between anode and cathode phenomena, the actual discharges used in the experimental studies with operating spark-machine tools have been of damped-oscillatory form so that each electrode serves as anode and cathode alternately, albeit at different current levels. Our investigation was accordingly concentrated on the effects of single, unidirectional time-limited discharges which would be expected to delineate physical differences at the anode and cathode more sharply.

The test procedure involved the initiation of sparkover between prepared anode and cathode electrodes at appropriate spacings by application of voltages of the order of 50 to 1,000 volts. The circuit was arranged to deliver, following sparkover, a single current pulse of substantially constant amplitude and controlled duration at convenient values in the range of 17 to 1,500 millimicroseconds. The circuit arrangements are shown in Fig. 1. The liquid dielectric employed was kerosene from a single bulk sample.

The electrodes used were annealed specimens of metals or alloys with the properties shown in Table I. Tensile strength and resistivity values were determined by test, while other properties listed are from standard references.

The craters produced by each single discharge were examined and optical measurements of depth and area were taken with calibrated microscopes. The effect on the quantitative data of statistical variations from crater to crater was minimized by averaging ten quantitative crater measurements for every electrical condition. The resultant data on average crater size, depth, etc., were readily repeatable and were found to vary consistently with electrical conditions, especially for the larger currents and current durations at which crater diameters and depths are large enough to be easily measured optically.

Fig. 2 shows photomicrographs of four typical craters in a cathode electrode produced by a short-duration discharge. The craters *C* and *D*, formed in a sintered-carbide cathode with a discharge using an anode of the same material, resemble those encountered in anodes; in particular *D* displays at its edges the outline of grain boundaries in the crystal structure, an effect often detected in anode craters. The craters *A* and *B* are those produced in an identical tungsten carbide cathode with an anode of brass. These suggest a relatively undisturbed surface "sprayed" with metal vapor or liquid metal which has subsequently solidified. This appearance is more or less characteristic of all cathode craters in hard materials when

---

Paper 57-2, recommended by the AIEE Dielectrics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted May 3, 1956; made available for printing October 17, 1956.

EVERARD M. WILLIAMS and RICHARD E. SMITH are with Carnegie Institute of Technology, Pittsburgh, Pa.

Work described in this paper was supported in part by the Method X Division of the Firth Sterling Corporation, Pittsburgh, Pa.

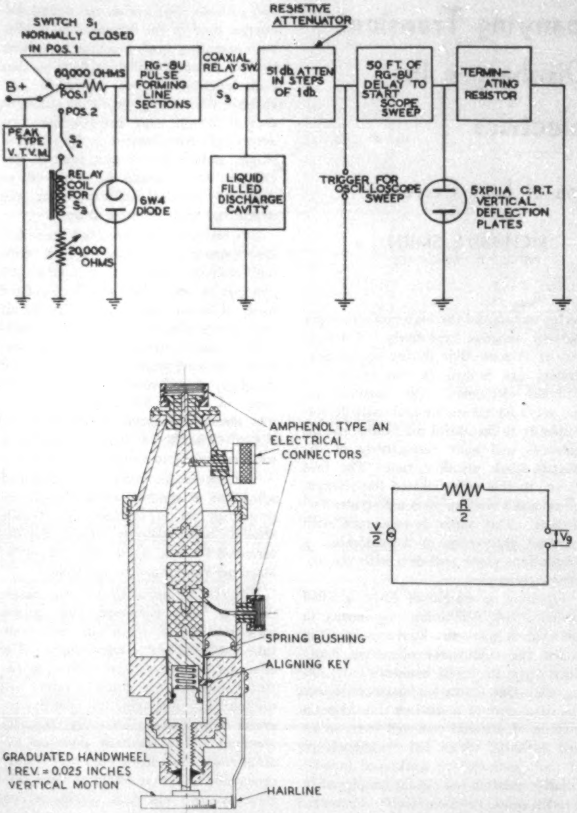

RESISTIVE ATTENUATOR
SWITCH S_1 NORMALLY CLOSED IN POS. 1
B+
60,000 OHMS
POS.1
POS. 2
RG-8U PULSE FORMING LINE SECTIONS
COAXIAL RELAY SW.
S_3
51 db. ATTEN IN STEPS OF 1 db.
50 FT. OF RG-8U DELAY TO START SCOPE SWEEP
TERMINATING RESISTOR
PEAK TYPE V.T.V.M.
S_2
RELAY COIL FOR S_3
6W4 DIODE
20,000 OHMS
LIQUID FILLED DISCHARGE CAVITY
TRIGGER FOR OSCILLOSCOPE SWEEP
5XP11A C.R.T. VERTICAL DEFLECTION PLATES
AMPHENOL TYPE AN ELECTRICAL CONNECTORS
SPRING BUSHING
ALIGNING KEY
GRADUATED HANDWHEEL 1 REV. = 0.025 INCHES VERTICAL MOTION
HAIRLINE
R/2
V/2
V_g

Fig. 1. Top—Circuit arrangements for experimental study of electrode craters. Closing of switch S_1 results in discharge of length of coaxial cable into liquid-filled discharge cavity. Pulse length determined by length of cable used. Discharge cavity, at lower left, together with equivalent circuit of source and load, at lower right

an anode of relatively soft material is employed. No analogous phenomenon occurs in craters formed in the anode since their form and depth are independent of the hardness, the melting point, or other physical characteristics of the cathode, except for cathodes of very low melting-point materials.

Although hard, strong anodes of high melting-point have been demonstrated to produce deep cathode craters, whereas soft anodes of low tensile strength and low melting point result in shallow cathode craters, a significant single characteristic or relation of anode to cathode characteristics has not been uncovered; this is unfortunate since a precise correlation would cast a powerful light on the exact mechanism of cathode-crater formation. This situation at the cathode differs markedly from that at the anode, inasmuch as anode crater depth is substantially independent of the cathode material. Moreover, anode data exhibit a simple correlation of anode crater depth with tensile strength consequently giving support to the electric field force theory of anode erosion.[3] In the absence of a comparable physical foundation, cathode data presented herein are organized along empirical lines.

The quantitative effect of anode material on cathode erosion is of considerable interest since cathode erosion corresponds

Table I. Table of Properties of Test Electrodes

Specimen	Alloy Designation	Resistivity $\times 10^{-6}$, Ohm-Cm*	Density, Pounds per Inch^3	Tensile Strength K, Pounds per Square Inch	Specific Heat, Cal/Gm/C†	Heat of Fusion, Cal/Gm	Heat of Vaporization, Cal/Gm	Melting Point, C	Boiling Point, C	Brinell Hardness
Tungsten	pure	5.56	0.690	597	0.032	61.0	960	3,370	6,000	290
Molybdenum	pure	5.70	0.325	270	0.061	70.0	1,460	2,620	4,700	250
18-8-SMO steel	316	76.7	0.29	125	0.12	69.8	1,528	1,385	2,830	160
Hot-rolled monel		54.0	0.318	96	0.13	68.0	1,313	1,330	2,830	162
FC steel	416	60.5	0.28	91	0.15	68.6	1,551	1,495	2,820	165
18-8-SM steel	303	76.8	0.29	81	0.12	70.0	1,530	1,425	2,820	140
17ST aluminum	17ST	5.56	0.102	64	0.23	91.5	2,513	645	2,490	110
Free cutting brass	240	7.95	0.307	61	0.09	39.6	931	885	1,910	131
Hard phosphor bronze B		10.10	0.321	60	0.09	46.7	1,210	1,050	2,490	128
Naval brass	386	7.23	0.304	75	0.09	39.8	928	900	1,860	136
Soft copper	204	1.76	0.323	32	0.092	48.9	1,250	1,087	2,500	42
Magnesium	pure	4.86	0.063	28	0.25	85.7	1,315	651	1,103	35
Zinc	pure	6.66	0.258	24	0.09	26.3	440	420	907	33
2S aluminum	2S	3.12	0.098	19	0.23	93.5	2,580	660	2,500	42
Cadmium	pure	7.59	0.312	12	0.055	13.6	212	320	768	33
Lead	pure	21.9	0.398	2.2	0.03	5.75	206	327	1,750	3.8
Tungsten carbide	HAX3		0.570	150				2,800		

* Centimeters.

† Calories per gram per C.

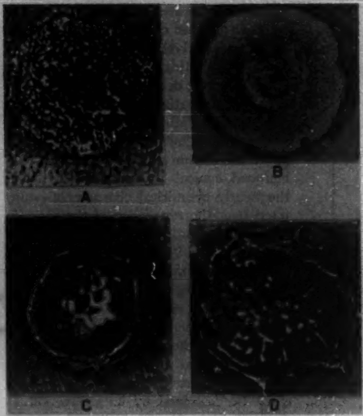


Fig. 2. Photomicrographs of some typical cathode craters produced by single discharge. A and B for discharge between brass anode and sintered tungsten-carbide (with cobalt binder) cathode. Crater A—11-ampere 0.3-microsecond pulse, actual crater diameter approximately 0.0002 inch; B—22.4-ampere 1.2-microsecond pulse, actual crater diameter approximately 0.0007 inch. C and D both anode and cathode of sintered tungsten carbide; C—11-ampere 0.3-microsecond pulse, actual crater diameter 0.0002 inch; D—11-ampere 1.2-microsecond pulse, actual crater diameter 0.0005 inch

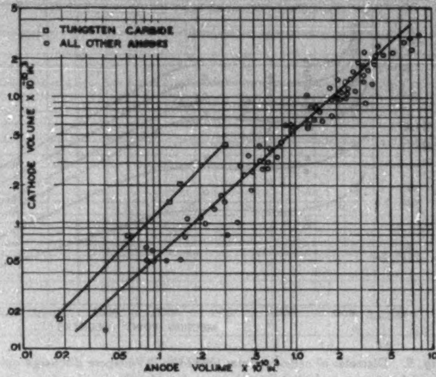


Fig. 3 (above right). Cathode-crater volume as function of anode-crater volume in craters produced by 11-ampere discharges (of wide range of durations) between anodes of various materials and brass cathode

to tool wear in the electric-spark machining process. Fig. 3 depicts this effect for the usual case of the brass tool-electrode, or brass cathode, with anodes of all materials tested. Except for the case of the extremely hard material, sintered tungsten-carbide, the anode crater size exceeds the cathode crater size, although by only a small ratio. The importance of utilizing the cathode as the tool, although apparent here, is more evident when the usual polarity is reversed; if a brass tool is used as an anode, to machine a tungsten-carbide workpiece (used as a cathode), tool wear rises to at least twenty times workpiece wear. The necessity of avoiding, or at least minimizing, oscillatory effects and corresponding "cutting" current reversals in the discharges is thus clear if tool wear is to be held to a minimum. Moreover, the impossibility of an improved spark-machine tool with the negligible wear of a soft, readily formed tool for machining very hard materials is shown.

Conclusions concerning the shape of the ionized path are based on the assumption that the area of a crater in either electrode under particular electrical conditions is identical with the area of the ionized path at the close of the discharge interval. Discharge areas are readily correlated with melting points of electrode materials. In temperature ranges in which discharge areas depend on melting temperature, the electrode material may be assumed to affect space-charge and possibly emission phenomena. Fig. 4 shows the relationship of cathode-crater diameter and the melting point of the anode for various discharge durations in an 11-ampere discharge. These curves are identical in shape with similar curves showing the relationship of anode-crater diameter and the melting point of the anode. The cathode craters are, however, slightly smaller than those at the anode. Both the anode craters and the cathode craters are independent in form and area of the melting point of the anode for a wide range of materials but, as Fig. 4 indicates, there are variations

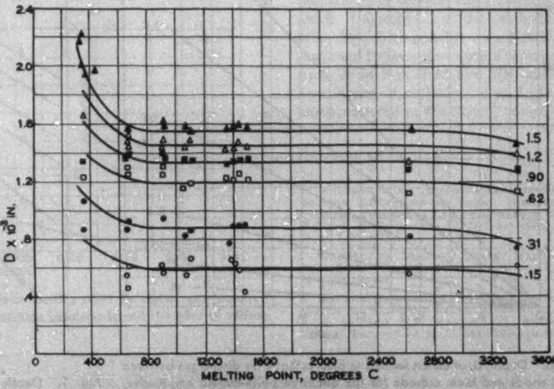


Fig. 4. Diameter of cathode crater formed by 11-ampere discharge pulses (with range of durations, in microseconds) between brass cathode and anodes of various materials plotted as function of anode melting-point

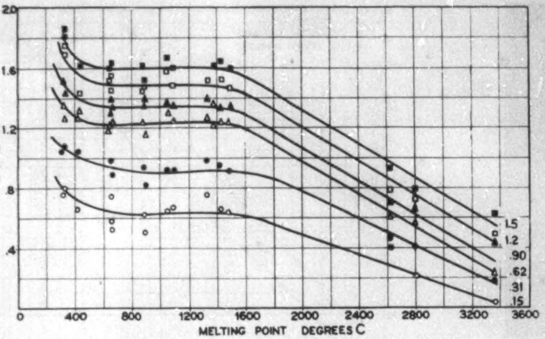


Fig. 5. Diameter of cathode crater formed by 11-ampere discharge pulses (with range of durations, in microseconds) between brass anode and cathodes of various materials plotted as function of cathode melting point

in the cathode crater size with anodes of very high melting point and those of very low melting point. The increase in diameter of the anode crater for low temperature anode materials has been attributed to the effect of metal-vapor ions on the space-charge neutralization mechanism and this effect is apparently communicated over the entire volume of the discharge to its terminus in the cathode surface.

Fig. 5 shows the relationship between the cathode-crater diameter and the melting point of the cathode. For cathodes of very low melting temperature the crater diameter increases markedly. A corresponding increase in crater diameter is, incidentally, communicated to the anode where the crater is again, however, slightly larger. For cathodes with melting temperatures from about 600 to 1,400 C (degrees centigrade) the cathode areas are relatively constant. Over the 1,400 C melting point the cathode-crater diameter decreases markedly. Under this latter cathode condition, however, anode-crater diameters remain constant so that for cathodes of very high melting point the discharge path assumes a pronounced conical character rather than the nearly cylindrical character encountered over the remainder of the cathode material range.

For the case in which both anode and cathode are formed of brass, some simple empirical equations for depth and volume of anode and cathode craters may be deduced. Figs. 6 and 7 show the dependence of the cathode crater depth on peak current and discharge duration. The cathode-crater depth d can apparently be described by an expression of the form.

$$d_c = K_d(I)^{1/4}(T)^{1/3}$$

in which d_c is depth in microns, K_d is a constant, T is discharge duration in microseconds and I is peak current in amperes. A graph of the volume of the crater produced at the cathode versus

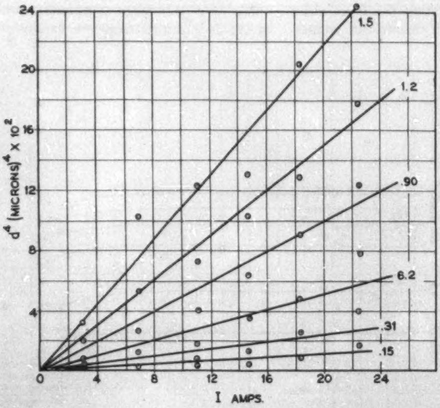


Fig. 6. Depth, d, of craters formed in brass cathode by discharges between brass anode and brass cathode for the discharge pulse-current amplitudes, in amperes, and durations, T, in microseconds, indicated. Depth plotted as function of d^4 to emphasize approximately linear relation between current and fourth power of depth

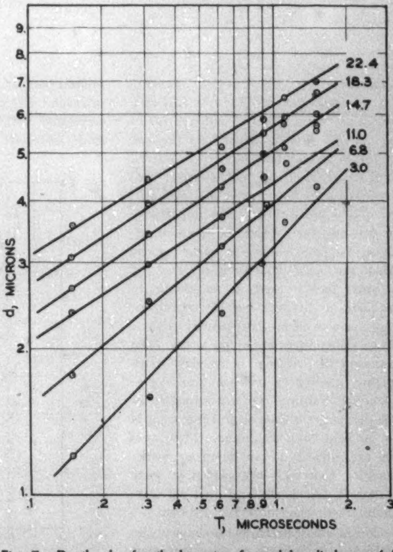


Fig. 7. Depth, d, of cathode craters formed by discharges (of various durations, T, in microseconds, and currents in amperes) between brass anode and brass cathode. Slope of upper curves is 0.32; thus, depth$\simeq K_d T^{1/3}$

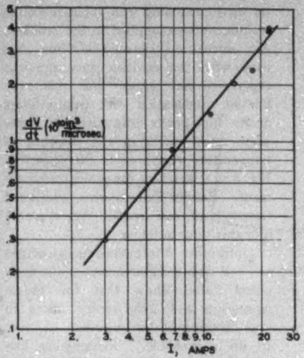


Fig. 8. Curve of derivative of cathode crater volume with respect to current duration, as function of current for discharges between brass anode and cathode. This curve can be drawn because volume varies linearly with duration. Slope of this curve is 1.23 so that

$$\frac{dV}{dT} \simeq K_v I^{1.25} \quad \text{and} \quad V \simeq K_v \times 10^{-9} I^{5/4} T$$

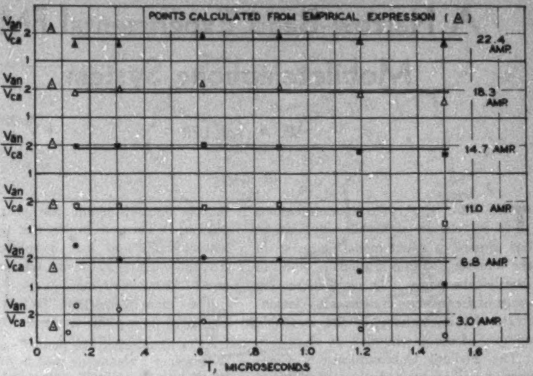


Fig. 9. Ratio of anode crater volume to cathode crater volume as function of current duration of various pulse currents with brass electrodes. Calculated points show values obtained from empirical expression

$$V_{anode}/V_{cathode} = I^{1/4}$$

pulse duration is a straight line. If the slope of this line is plotted against peak current, curves such as Fig. 8 are obtained. From these results for brass electrodes, an empirical expression for the crater volume V, in cubic inches, is

$$V_c = 0.0075 \times 10^{-9} T(I)^{5/4}$$

The corresponding expression for the brass anode has been shown[1] to be

$$V_a = 0.0075 \times 10^{-9} T(I)^{3/2}$$

A comparison of the ratio of anode to cathode crater volumes

$$V_{anode}/V_{cathode} = I^{1/4}$$

as determined from these expressions, and those actually measured under a variety of electrical conditions, is shown in Fig. 9.

In spite of the relatively simple relations disclosed for cathode craters, the authors have found no such satisfactory basis for a quantitative theory of cathode erosion as that determined earlier for anode erosion.[1] Cathode-crater volume depends in a general way upon the melting point of the cathode material, as shown in Fig. 10; this effect and the direct relation of cathode volume to discharge duration suggest that purely thermal-energy effects are involved. Moreover, substitution of a ferromagnetic for a nonferromagnetic cathode material of approximately like melting point, results in no change in crater depth, and indicates therefore that the energy for melting arises from the discharge rather than from the energy dissipated by current in the surface of the cathode. A theory of cathode erosion should certainly include, however, the mechanical or thermal effect of impact of anode fragments on the cathode in solid, liquid, or possibly vapor state, as suggested by the appearance of superposed particles in Figs. 2(A) and 2(B) and the dependence of cathode crater depth on anode material in the "hard" range.

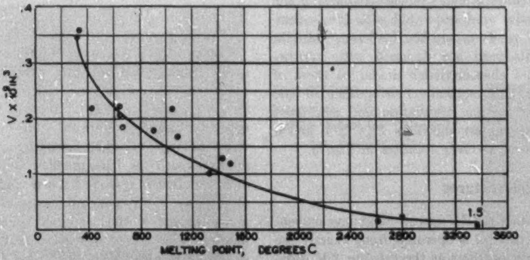


Fig. 10. Cathode crater volume as function of cathode melting point for craters produced by 11-ampere discharge of 1.5 microseconds duration between brass anode and cathodes of various materials

References

1. Phenomena Accompanying Transient Low-Voltage Discharges in Liquid Dielectrics. I—Anode Phenomena at Low Currents, E. M. Williams, R. E. Smith. *AIEE Transactions*, vol. 74, pt. I, May 1955, pp. 164–69.

2. Theory of Electric Spark Machining, Everard M. Williams. *Ibid.*, vol. 71, pt. II, Jan. 1952, pp. 105–08.

3. Recent Developments in the Theory and Design of Electric Spark Machine Tools, E. M. Williams, J. B. Woodford, Jr., Richard E. Smith. *Ibid.*, vol. 73, pt. II, May 1954, pp. 83–88.

A Narrow-Band Experimental FM Mobiletelephone System

W. A. MILLER
MEMBER AIEE

THE VALUE of radio communications in all classes of mobile applications has caused a phenomenal increase in the number of radio stations licensed in these vital services. This ever-increasing usage of frequencies continues to demonstrate the inadequacy of the available number of usable channels.

The Federal Communications Commission had recognized the need for more channels and issued proposed rules early in 1955 under Docket 11253. The proposals were based on creating more channels by reducing channel spacing (splitting channels) in certain frequencies allocated to the mobile services.*

Manufacturers, several years ago, demonstrated that radio equipment of latest design will readily permit mobile radio communications using narrow-band principles in which the transmissions occupy only one half of the present assigned channel width. The most recent improvements in transmitter frequency stability and receiver selectivity further reduce the possibility of interference between closely spaced (split-channel) mobile radio systems.

The performance of a narrow-band channel in the domestic public mobile telephone service had not been tried and only a limited amount of study data was available. To test the accuracy of study data under actual operating conditions an experimental narrow-band channel was planned with the use of 30 to 44-mc frequencies. This frequency band was chosen for a trial because channel assignments in these mobile telephone services are made according to a nation-wide zone plan. Present and proposed frequency distribution are given in the zone plan in Fig. 1.

The trial channel has been operating in Los Angeles about one year. The station is licensed under the developmental rules and uses a split-channel frequency pair, 35.52 mc transmitting, and 43.52 mc receiving.

The more important factors being evaluated during this trial are:

1. The adequacy of the present zone plan of frequency assignments used in this service if channels are split and suitably distributed among the zones.
2. Impairment to transmission if adjacent zone assignments and assignments to zones 600 to 1,500 miles apart were reduced to 20 kc during periods of appreciable *E*-layer transmissions.
3. Impairment to transmission if channel spacing in the same zone and in zones 2,000 miles or more apart were reduced to 20 kc during periods of appreciable *F*-layer transmissions.
4. Maintenance variables and their probable penalties on transmission.

Design of the narrow-band channel follows closely that of an existing regular 40-kc channel now serving the Los Angeles area on a frequency pair of 35.38 and 43.38 mc. Transmitter and receivers of the trial channel are installed in the same locations as those of the regular channel. The trial base station is a narrow-band transmitter-receiver combination. There are six satellite receivers in various parts of the service area associated with the system. All receivers are narrow-band. Mobile units used are designed with narrow-band characteristics similar to those of the base station. Photographs of the base station installation and associated antennas are shown in Figs. 2–4, and a typical receiver location in Fig. 5.

Trial Features

Since the *ZY* narrow-band developmental channel is operated from exactly the same locations as the existing *ZM* wide-band channel, several ideas were incorporated in the design. The two 250-watt base transmitters are connected to a single 6-db (horizontal gain) antenna of local design. Resonant cavity filters (two per transmitter) are used in the respective transmitter outputs. Insertion loss is one db per cavity. Quarter-wave matching and interconnection techniques were applied to the problem. Air dielectric transmission line was used with the cavities since previous experience has shown that this type line in combination with cavity filters and a high-power transmitter is somewhat more stable in the presence of temperature variations in the dielectric. Fig. 6 is a sketch of the present arrangement. Results indicate that additional transmitters can readily be operated into this same antenna.

Third-order intermodulation products are well within requirements. The computed values show that for 140-kc separation and 2-db cavity filters in each transmitter a rejection of over 80 db is possible. Recognizing loss through the cavities and the antenna gain, the effective radiated power is 600 watts on the regular channel and 750 watts on the narrow channel. A substantial portion of this difference is traceable to the higher efficiency of the newer trial system transmitter.

At receiving locations, one wide-band and one narrow-band receiver are operated from a common antenna using a bridging arrangement made of solid dielectric transmission lines. Measured bridging loss is 2.8 db. At one location an electronic antenna matching unit is being tried. The measured gain of the matching unit is 4.5 db per receiver. These matching units were proposed at all receiving points if additional narrow-band channels might be established. Consideration has been given to applying an off-channel squelch circuit to receivers if nuisance interference becomes troublesome.

System Line-Up

Transmission levels in the experimental radio channel are set to establish a maximum modulation deviation of ±7 kc as compared to ±15 kc used in the regular channel. With a 1,000-cycle test tone at +7 decibels referred to one milliwatt into the base transmitter, the deviation limiter was set for a deviation of ±7 kc. Subsequently, with 1,000 cycles at zero decibels referred to one milliwatt, the transmitter input control was set for a deviation of 2.5 kc (the reduced level allows a peak factor of 7 db). When telephone circuits are connected to the radio channel, it is necessary to control the wide variation in speech levels and talker volumes encountered in these circuits. This is accomplished by an automatic gain-adjusting amplifier in the radio control

---

Paper 57-58, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 19, 1956; made available for printing December 28, 1956.

W. A. Miller is with The Pacific Telephone and Telegraph Company, Los Angeles, Calif.

---

*On September 24, 1956, the Commission released public notice 56-901 adopting rules relating to split-channel assignments under Docket 11253. The 30- to 44-mc (megacycle) band was not included in the rules thus amended.

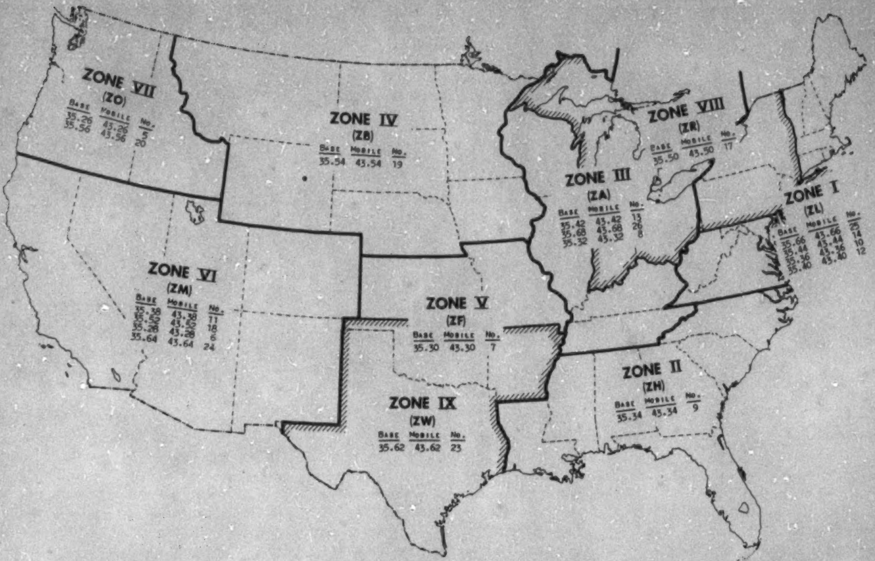

ZONE VII
(ZO)
Base 35.26 35.56
Mobile 43.26 43.56
No. 5 20
ZONE IV
(ZB)
Base 35.54
Mobile 43.54
No. 19
ZONE VI
(ZM)
Base 35.38 35.52 35.28 35.64
Mobile 43.38 43.52 43.28 43.64
No. 11 18 6 24
ZONE V
(ZF)
Base 35.30
Mobile 43.30
No. 7
ZONE IX
(ZW)
Base 35.62
Mobile 43.62
No. 23
ZONE III
(ZA)
Base 35.42 35.68 35.32
Mobile 43.42 43.68 43.32
No. 13 26 8
ZONE VIII
(ZR)
Base 35.50
Mobile 43.50
No. 17
ZONE I
(ZL)
Base 35.66 35.44 35.36 35.40
Mobile 43.66 43.44 43.36 43.40
No. 25 14 10 12
ZONE II
(ZH)
Base 35.34
Mobile 43.34
No. 9

Fig. 1. Mobile telephone service; 30- to 44-mc band. Proposed frequency allocation plan based on 20-kc channel spacing

terminal where speech level variations are held to about 3 db. Input levels of the 600- and 1,500-cycle dial signal tones are adjusted to produce a transmitter deviation of 3 kc and 6 kc respectively. This allows 1 kc for variables.

Comparable deviation adjustments were made in the mobile units so as to assure that ±7-kc deviation would rarely be exceeded for any talker level condition. A frequency meter having an accuracy of ±0.0005 per cent is used for narrow-band frequency checking and modulation deviation measurements. Since the frequency tolerance of the transmitters and receivers used is specified at 0.003 per cent, the use of a highly accurate checking device tends to minimize transmission penalties likely to arise from small adjustment errors.

Experience with the trial channel indicates that performance of receivers used in this narrow-band mobile telephone service appears to be free of objectionable distortion when carriers modulated up to about ±9.5 kc are being received. In the case of the very latest receivers, distortion begins to appear when carriers having ±8-kc deviation are received. Four mobile units equipped with $+7\frac{1}{2}$-kc receivers have recently been added to the system.

Fig. 2 (left). Base station transmitters, 30- to 44-mc mobile telephone channels, Los Angeles area

Fig. 3 (right). Base station location. Los Angeles area mobile telephone system. Mt. Verdugo, elevation 3,126 feet

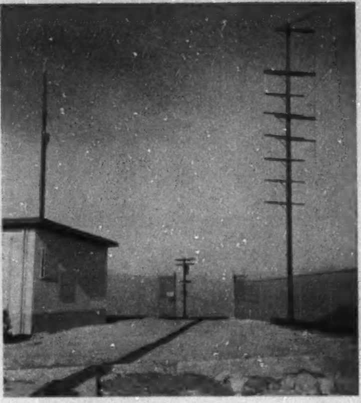


Fig. 4 (left). Gain antenna, 30- to 44-mc mobile telephone channels. Los Angeles area

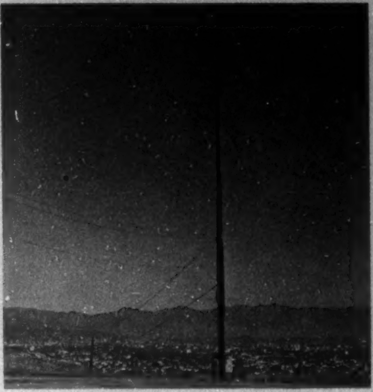


Fig. 5 (right). Baldwin Hills receiver site. Los Angeles area mobile telephone system

Interference

In the Los Angeles area a paging service operated on channel 21, 43.58 mc, with a 250-watt amplitude-modulated transmitter. The station is less than 1/2 mile from a receiver associated with the developmental mobile channel. Spurious signals from the paging station transmitter (60 kc separated) caused excessive off-channel nuisance interference in the receiver. When desired mobile signals are received at this receiver the noise degradation was approximately 15 db or about one-circuit merit. Work with the licensee in improving frequency stability and amplitude modulation has reduced interference to a negligible value.

Adjacent channel interference due to sky wave propagation has not been experienced in the narrow-band receivers during the one year of trial channel operation. Recent observations were made to determine the amount of adjacent channel signals reaching the Los Angeles area due to *E*- and *F*-layer propagation from eastern zones. Receivers, equipped with limiter current recorders and tuned to the adjacent channels, 20 kc on each side of the experimental channel, were monitored for a period of about 4 weeks during September 1956. The lower adjacent frequency, 35.50 mc used in zone VIII, 2,000 miles or more distant, was received with varying values of signal, none of which exceeded 70 microvolts, the average being about 45 microvolts (dipole receiving antennas). The upper adjacent frequency, 35.54 mc used in zone IV, 1,500–2,000 miles distant, was received in varying values of signal none of which exceeded 50 microvolts, the average being about 20 microvolts. The higher degree of selectivity in the newer receivers was further emphasized during preliminary monitoring tests which showed that the $7\frac{1}{2}$-kc receiver operating 20 kc from the experimental channel was not disturbed when the experimental base transmitter was radiating full power. Under this same condition a receiver of an earlier model was subject to off-channel nuisance interference and a noise penalty of 3 to 5 db. The adjacent channel monitor receivers were 10 miles by airline from the Los Angeles experimental transmitter.

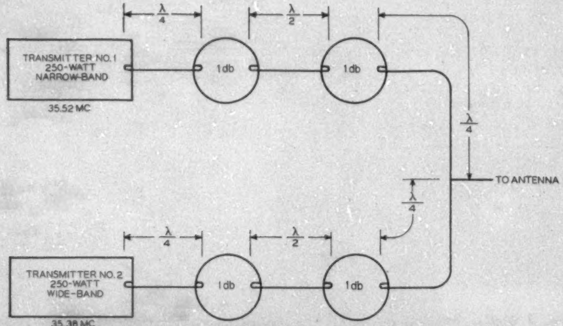

λ/4
λ/2
TRANSMITTER NO.1 250-WATT NARROW-BAND
35.52 MC
1db
1db
λ/4
TO ANTENNA
λ/4
λ/4
λ/2
TRANSMITTER NO.2 250-WATT WIDE-BAND
35.38 MC
1db
1db

Fig. 6. Duplex connections: two transmitters to common antenna

P-8290 cavity resonators; 3/8-inch air dielectric line; constant impedance connectors

Service Area Coverage

Initial observations in the service area indicate that within a radius of 35 to 40 miles the coverage obtained from either of the channels is substantially equal; in fringe areas they indicate that the wide-band channel provides a slightly better speech-to-noise ratio.

Conclusions

Progress of the narrow-band trial in Los Angeles seems to verify predictions that mobile telephone service in the 30- to 44-mc band can readily be expanded on a reduced channel space basis. From the actual experience of use it appears that the probable transmission penalties of from 3 to 6 db and the need for more careful maintenance attention constitute a reasonable cost for obtaining 18 instead of 9 channels for distribution to the zones, especially where demand is heavy.

Elimination of the Ballast Resistor in D-C Power Supplies

STEWART M. KRAKAUER
NONMEMBER AIEE

BALLAST RESISTORS, when required with conventional d-c power supplies, can be eliminated by designing appropriate terminal load characteristics directly into "virtual-ballast" power supplies. There are several ways in which this might be done. One method, the inductive-ballast power supply, is presented in this paper. It permits operation at close to 100-per-cent efficiency and, moreover, places minimal stress on circuit components. This type of supply may therefore be considerably smaller, less expensive, and more efficient than its conventional equivalent.

The Ballast Resistor

Standard practice in the design of power supplies for d-c arcs calls for a ballast resistor in series with the arc.[1] Factors that enter into the choice of this resistance value are

1. Resistance of the series combination of ballast resistor and incremental resistance of the arc must be positive for stable operation.

2. High values of resistance lead to less variation of arc power with respect to fluctuation in the a-c line voltage. Greater arc stability and protection against extinction therefore result.

3. The arc must be adequately accommodated through its starting transient to proper steady-state operation.

4. Excessive waste of power and stress upon the power-supply components should be avoided.

Conditions 1 and 2 set a lower limit on this resistance value, condition 3 sets a range, and condition 4, an upper limit. The ballast resistance may be nonlinear. If so, condition 1 is interpreted to refer to incremental resistance near the operating point, conditions 2 and 3 to a locus, and condition 4 to the d-c resistance at the operating point. Power dissipated in the ballast resistor is usually the same order of magnitude as that in the arc.

Ballast resistors are also required in those other circuits where normal changes in load resistance would otherwise lead to excessive variation in load power. Heaters, for instance, usually have a much lower resistance when cold than when they are at normal operating temperature. Sudden application of full operating voltage therefore results in a high-power starting transient with possible consequent damage to the heater or to the supply. In other situations, wide swings in load resistance during operation might have to be accommodated, or it may be required to establish specified variation in load power with load resistance.

A design technique is treated here which does away with the ballast resistor altogether. The resulting power supply may operate at a much higher efficiency, and impose much less stringent demands upon its components than would be possible by conventional means.

The Virtual-Ballast Power Supply

A ballast resistor is needed with the usual constant-voltage power supply in order to impose appropriate volt-ampere relationships on the load. The question arises as to whether an appropriate load line could be provided by some type of power supply acting alone, i.e., a virtual-ballast power supply.

It is important to keep in mind that similarity of two circuits on the basis of their load lines does not at all necessitate that their transient impedances be similar. Consideration of transient impedances is particularly important when the load is an arc, for if it is too low instability and extinction may result. Accurate evaluation of this impedance may be very difficult, but it can be conveniently approximated by considering output impedance of the filter network alone. More accurate knowledge of output impedance may be of little practical value because so little will probably be known of the transient impedance of the associated arc. A series inductor is better at the filter output than a shunt capacitor because output impedance may then be kept sufficiently high. This inductor has the additional advantage of offering a convenient means for starting the arc if its required trigger voltage exceeds that of the supply at open circuit.[1]

Performance characterized by a virtual-ballast resistor does not necessarily imply that there is any actual power dissipation associated with the ballast action. High operating efficiencies are therefore attainable.

Design of a virtual-ballast power supply might be approached in any of several ways. Conventional voltage-multiplier circuits would be appropriate in some cases. Regulation in these circuits is subject to control through choice of circuit configuration and through choice of capacitance values. Load characteristics for some of these circuits are available in the manufacturer's literature[2,3] and will not be detailed here.

Requirements might justify coupling the load, through rectifiers and filters, to a high-frequency oscillator. This approach would be particularly appropriate when high stability and/or low-power operation is required.

Another approach involves use of a conventional power supply with a reactive element on the a-c side of its rectifier. The remainder of this paper will be devoted to discussion of a particular circuit which stems from this last approach. The inductive-ballast power supply appears to have some useful and interesting properties.

The Inductive-Ballast Power Supply

A-C BALLAST

It will be useful for purposes of orientation and comparison to find load lines for the a-c circuit shown in Fig. 1. This circuit might arise from the use of a physical series inductive ballast or from consideration of the equivalent circuit for a ballast transformer. A purely resistive load is assumed. Currents and voltages are sinusoidal and are considered in terms of their rms values.

$$E_o = E_a R_o / |R_o + \bar{Z}_g| \quad \text{where } \bar{Z}_g = R_g + j\omega L_g$$
$$I_o = E_a / |R_o + \bar{Z}_g| \qquad (1)$$

It will be convenient to express these equations in normalized form; the coordinates being taken as: E_o/E_a, I_o/I_s, R_o/Z_g, Q_g where $I_s = E_a/Z_g$, and $Q_g = \omega L_g/R_g$. Voltage is normalized with respect to its open-circuit value, and current with respect to its short-circuit value. The two system parameters R_o/Z_g and Q_g then determine the response parameters of

---

Paper 57-78, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 22, 1955; made available for printing December 17, 1956.

STEWART M. KRAKAUER is with The United States Time Corporation, Irvington-on-Hudson, N. Y.

This paper was based on work performed under contract *AT-30-1-GEN-70* for the Atomic Energy Commission at Radiological Research Laboratory, Columbia University.

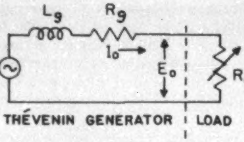


Fig. 1. A-c inductive-ballast schematic

the system E_o/E_a and I_o/I_s as follows:

$$\frac{E_o}{E_a}=\frac{1}{\left[1+2\frac{Z_g}{R_o}\frac{1}{\sqrt{Q_g^2+1}}+\left(\frac{Z_g}{R_o}\right)^2\right]^{\frac{1}{2}}}$$

$$\frac{I_o}{I_s}=\frac{1}{\left[1+2\frac{R_o}{Z_g}\frac{1}{\sqrt{Q_g^2+1}}+\left(\frac{R_o}{Z_g}\right)^2\right]^{\frac{1}{2}}} \qquad (2)$$

Equation 2 is plotted in Fig. 2. Dependence of the load line shape upon Q_g which characterizes the source impedance, and location of the operating point depending upon the ratio R_o/Z_g is clearly shown. If a series capacitor were used instead of the inductor, similar curves would result.

The point should perhaps be made to avoid confusion, that the foregoing analysis is not intended to depict load-line characteristics for a ballast transformer operating into an arc. The arc impedance is far from being a pure resistance, and the wave forms involved are not sinusoidal.

UNFILTERED INDUCTIVE-BALLAST POWER SUPPLY

Suppose that a rectifier is interposed in the R-L circuit between the generator and load. This circuit will be called an unfiltered inductive-ballast power supply.

A bridge-type rectifier is taken for concreteness and it will be assumed that the rectifying elements may be adequately represented by a resistor R_c when voltage polarity is in the conducting direction, and R_n when nonconducting. The unfiltered inductive-ballast power-supply circuit is shown in Fig. 3(A) and its equivalent in Fig. 3(B). For convenience, the bridge-resistance network may be converted to its equivalent T as shown in Fig. 3(C).[4] If attention is directed to the rms volt-ampere characteristics at the load, the existence of the ideal rectifying elements, as shown in Fig. 3(D), may be ignored since they do not influence rms load conditions at all.

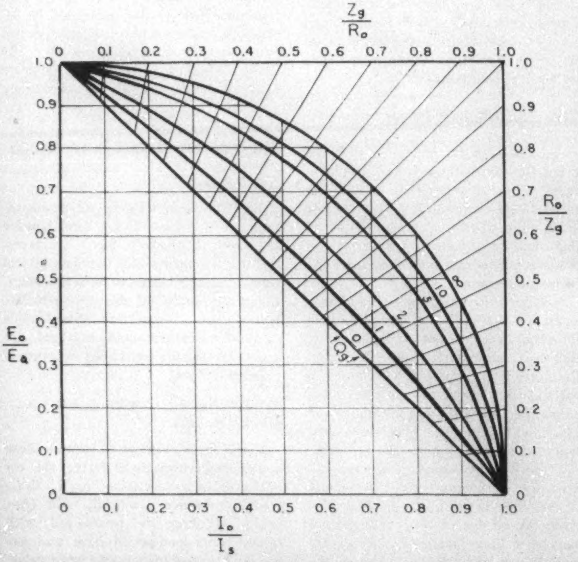


Fig. 2. Normalized load lines for the a-c inductive-ballast circuit and for the lossless unfiltered inductive-ballast power supply

Equations may now be set up from Fig. 3(D).

$$\begin{cases} E_a=\left(\bar{Z}_g+\dfrac{R_n+R_c}{2}\right)\bar{I}_1-\left(\dfrac{R_n-R_c}{2}\right)\bar{I}_o \\ 0=-\left(\dfrac{R_n-R_c}{2}\right)\bar{I}_1+\left(R_o+\dfrac{R_n+R_c}{2}\right)\bar{I}_o \end{cases} \qquad (3)$$

Normalization now requires two additional parameters to account for the lossy rectifier. These are taken as R_g/R_c and R_c/R_n.

Solving equation 3 leads to

$$\frac{E_o}{E_a}=\frac{1}{\sqrt{A^2+B^2}}$$

$$\frac{I_o}{I_s}=\left(\frac{E_o}{E_a}\right)\left(\frac{Z_g}{R_o}\right) \qquad (4)$$

where

$$A=\left(\frac{Z_g}{R_o}\right)\left(\frac{1}{\sqrt{Q_g^2+1}}\right)\left(\frac{1+\frac{R_c}{R_n}+2\frac{R_c}{R_g}}{1-\frac{R_c}{R_n}}\right)+\left(\frac{1+\frac{R_c}{R_n}\left(1+2\frac{R_g}{R_c}\right)}{1-\frac{R_c}{R_n}}\right)$$

$$B=\left(\frac{Z_g}{R_o}\right)\left(\frac{Q_g}{\sqrt{Q_g^2+1}}\right)\left(\frac{1+\frac{R_c}{R_n}}{1-\frac{R_c}{R_n}}\right)+\frac{2\left(\frac{R_g}{R_c}\right)\left(\frac{R_c}{R_n}\right)}{1-\frac{R_c}{R_n}}$$

Graphical representation of equation 4 cannot be made as concisely as was done for equation 2 because of the increased number of variables involved. A convenient measure for the effect of rectifier losses can be had by considering the shift in open- and short-circuit characteristics due to these losses.

$$\left.\frac{E_o}{E_a}\right|_{\frac{R_o}{Z_g}\to\infty}=\frac{1-\frac{R_c}{R_n}}{\left[\left(1+\frac{R_c}{R_n}+2\frac{R_g}{R_n}\right)^2+\left(2\frac{R_g}{R_n}\right)^2\right]^{\frac{1}{2}}} \qquad (5)$$

$$\left.\frac{I_o}{I_s}\right|_{\frac{R_o}{Z_g}\to 0}=\left(1-\frac{R_c}{R_n}\right)\times \left[\frac{Q_g^2+1}{\left(1+\frac{R_c}{R_n}\right)^2(Q_g^2+1)+\left(1+\frac{R_c}{R_n}+\frac{R_c}{R_g}\right)\left(4\frac{R_c}{R_g}\right)}\right]^{\frac{1}{2}}$$

It should be kept in mind that in equation 4, terminal voltage is normalized not with respect to the open-circuit terminal voltage of the existing circuit, but rather relative to the Thévenin generator source voltage. Similarly, the load-current reference is that due to the short-circuited Thévenin generator and not the existing network.

For the ideal rectifier ($R_c=0$, $R_n=\infty$) both expressions in equation 5 reduce to unity, and the resulting curves are identical to those shown in Fig. 2.

FILTERED INDUCTIVE-BALLAST POWER SUPPLY

Modes

If a filter is now interposed between the load and the rectifier in the previous circuit, a filtered inductive-ballast power supply is provided. As a first step in its analysis, it will be convenient to examine the modes of operation for this circuit. This may be done by considering equilibrium conditions in its lossless, constant parameter, infinite filter idealization. The characteristic behavior of this circuit forms an interesting contrast with a similarly idealized choke-input filter power supply. Bridge rectifiers are assumed for concreteness, although other types are applicable.

The filtered inductive-ballast power supply might be seen as a conventional choke-input filter power supply, but with its input choke shifted to the a-c side of the rectifier. These two circuits are shown in Fig. 4 together with their associated current and voltage wave forms. In accordance with the foregoing assumptions: constant voltage exists across C_1, L_1 has no associated resistance, the rectifiers act like switches, and the source impedance (with transformer included) is zero.

Let us briefly recall operation of the conventional choke-input filter power supply.[5] Assuming continuous rectifier conduction, voltage input to the filter (e_2) must be a full-wave rectified sinusoid. Current through $L_1(i_2)$ consists of a constant and a varying component, which are best considered independently. The d-c component ($\bar{i}_2$) is simply the load current (I_{DC}) and is determined by the value of R_o. The varying component ($\tilde{i}_2$) is determined by the voltage drop across L_1.

$$e_2-e_3=L_1(d\tilde{i}_2/dt)$$

Integrating and solving:

$$\tilde{i}_2=(E_1/\omega L_1)(1-\cos\omega t-2\omega t/\pi) \qquad (6)$$

where:

$$0<\omega t<\pi$$

and for the d-c component:

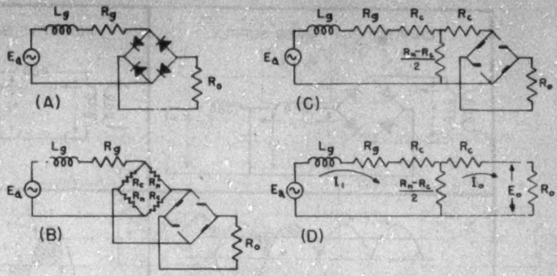


Fig. 3. Unfiltered inductive ballast power supply

A—Circuit schematic
B, C, D—Successive steps in the derivation of its equivalent circuit

$$\bar{i}_2=(2/\pi)(E_1/R_o)$$

where $e_1=E_1\sin\omega t$. Also

$$i_2=\tilde{i}_2+\bar{i}_2$$

The assumption of continuous rectifier conduction will hold so long as i_2 remains positive, which means that $\bar{i}_2$ must exceed the peak negative value of $\tilde{i}_2$. Expressed in more convenient terms; $R_0/\omega L\leqslant\sim 3$ for continuous conduction.

A small increase in R_o above this critical value causes interruption of the current, but commutation is maintained until further increase of R_o. The conduction angle remains at π up to the critical point and then decreases toward zero at open circuit.

The shaded areas in the voltage waves mark off conduction intervals. From the foregoing equation 6, $\tilde{i}_2$ is a function of the integral of this area. Since $\tilde{i}_2$ in each case returns to its initial value at the end of conduction, the upper and lower portions of the shaded area must be equal to each other.

Load characteristics of the conventional choke-input filter supply can be sketched from the foregoing analysis. For load currents exceeding that produced by the critical value of R_o terminal voltage is constant and equal to the average value of the applied rectified sinusoidal voltage. For decreasing currents, voltage increases smoothly until at no load, the direct voltage is equal to the peak value of the applied alternating voltage.

Analysis of the operating modes for the filtered inductive-ballast power supply begins with the same assumption of continuous conduction. This necessitates that a square-wave voltage appear at e_3. Current through L_1 alternates and must be passing through zero at the instants of e_3 voltage transition. This follows from consideration that current into the rectifier must always be "in phase" with rectifier voltage. Moreover, since the current must return to zero at the end of each conduction phase, it follows that the shaded areas are again equal, top and bottom.

Current through L_1 (i_1) is again determined from the known voltage drop across it. In this case there is no direct component of current, though when i_1 is rectified (i_2) a d-c value exists which is equal to the load current (I_{DC}). This circuit finds its own equilibrium by varying E_{DC} and ψ so that all the other restraints are satisfied.

When $R_o=0$, $E_{DC}=0$ and $\psi=\pi/2$. Increasing R_o causes an increase in E_{DC} and a decrease in ψ until the corner of e_3 touches the sinusoid e_1, Figs. 4(F) and (G). This condition defines the critical value of R_o for continuous conduction. Further increase of R_o causes the voltage at the load side of the rectifier (e_3) to exceed the magnitude of e_1 when the preceding cycle ceases conduction. Conduction cannot resume until the magnitude of e_1 again exceeds e_3. As R_o increases further, ψ increases again back to $\pi/2$ at open circuit, at which time E_{DC} is equal to the peak value of applied voltage e_1. In this circuit, therefore, terminal voltage may range from zero at short circuit to a maximum at no load.

Analysis

Analysis of the inductive-ballast power supply is based on the foregoing assumptions and two others. For convenience, they are all stated explicitly here. Necessity for one additional assumption will arise in the course of the development of the analysis.

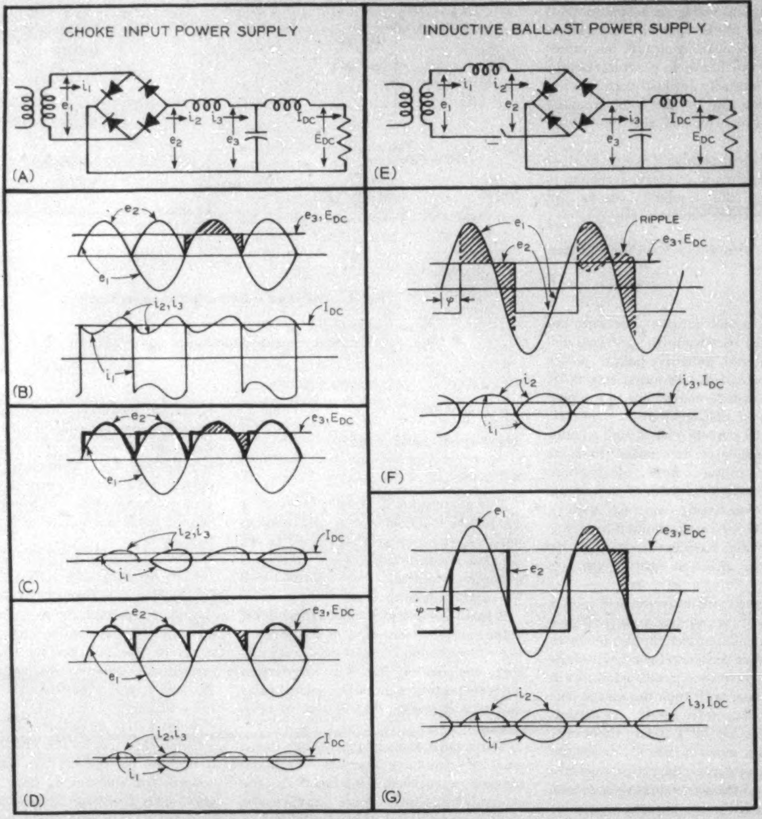


Fig. 4. Modes of operation for an idealized conventional choke-input filter power supply (A-D) and for an idealized filtered inductive-ballast power supply (E-G)

A—Choke-input power supply
B—Continuous conduction for A
C—Intermittent conduction, commutating, for A
D—Intermittent conduction, noncommutating, for A
E—Inductive-ballast power-supply
F—Continuous conduction for E
G—Intermittent conduction for E

1. The equivalent Thévenin generator for a high-reactance transformer, or for a transformer or line with an inductor in series, is adequately represented as a fixed sinusoidal voltage with fixed series resistance and inductance.

2. The rectifier elements may be considered to have a fixed (current-independent) resistance when the voltage polarity across them favors conduction, and a different fixed resistance when the polarity across them is reversed.

3. The voltage across the output terminals of the rectifier (input to the filter) is taken as constant.

4. The filter elements are dissipationless.

Discussion of the validity and influence of these assumptions is postponed until the results based on them are available. It will be seen that they lead to useful solutions.

One is led to expect, from consideration of the mode of operation for the ideal circuit, that there are cyclic intervals of conduction and nonconduction. Different equivalent circuits exist for the network during these periods as is shown in Fig. 5. A conflict now arises when transition from one state to the other is considered. Assume that the rectifiers are conducting so that Fig. 5(D) describes the prevailing conditions. When i_2 drops to zero the circuit cannot immediately switch to the nonconducting state shown in Fig. 5(H) because inductor L_g prevents any sudden change in i_1, e_b would then initially exceed E_{DC} which is contrary to the condition for existence of equivalent circuit 5(H). This conflict appears again in the transition back to conduction.

In actual practice the two states are not so sharply distinct because in the region of zero voltage there is a much smoother transition in resistance with respect to changing polarity than appears in the idealization of the rectifier. However, it was observed that in actual operation, an oscillation in voltage e_b appears to be associated with transition. Further treatment of this effect will not be undertaken here since it would complicate matters far out of proportion to any resulting gain in accuracy.

In order to proceed on the basis of linear parameters, it will be necessary to ignore the foregoing contradiction. It seems best to ignore the condition requiring continuity in i_1. The current increment involved is relatively small, and this assumption seems to do least violence in terms of its ultimate consequences.

Typical wave forms for this circuit which show also the nature of the preceding adjustment are shown in Fig. 6. The equations and their associated boundary conditions which define the behavior of this system are as follows:

During conduction:

$$\begin{cases} E_1 \sin(\omega t+\psi) = \left[L_g \dfrac{d}{dt} + R_g + \dfrac{R_n+R_c}{2}\right] i_1(t) - \left[\dfrac{R_n-R_c}{2}\right] i_2(t) \\ -E_{DC} = -\left[\dfrac{R_n-R_c}{2}\right] i_1(t) + \left[\dfrac{R_n+R_c}{2}\right] i_2(t) \end{cases} \quad (7)$$

with boundary conditions:

$$i_1(0) = i_1\left(\frac{\alpha}{\omega}\right) = \frac{2E_{DC}}{R_n-R_c}$$

During nonconduction:

$$E_1 \sin(\omega\tau+[\psi+\alpha]) = \left[L_g \frac{d}{d\tau} + R_g + R_n\right] i_1(\tau) \quad (8)$$

with boundary conditions:

$$i_1(0) = +\frac{E_{dc}}{R_n};\; i_1\left(\frac{\pi-\alpha}{\omega}\right) = -\frac{E_{DC}}{R_n}$$

And finally, the rectified average value of $i_2(t)$ must be equal to the load current plus current drained back through the rectifier during nonconduction.

$$I_{Dc} = \frac{\omega}{\pi}\int_0^{\alpha/\omega} i_2(t)\,dt - \left(\frac{E_{DC}}{R_n}\right)\left(\frac{\pi-\alpha}{\pi}\right) \quad (9)$$

The three dependent variables which most conveniently define the equilibrium state of this system are: E_{DC}/E_a, ψ, and α. Each of the foregoing equations implicitly expresses an independent relationship among these variables, so a set of three

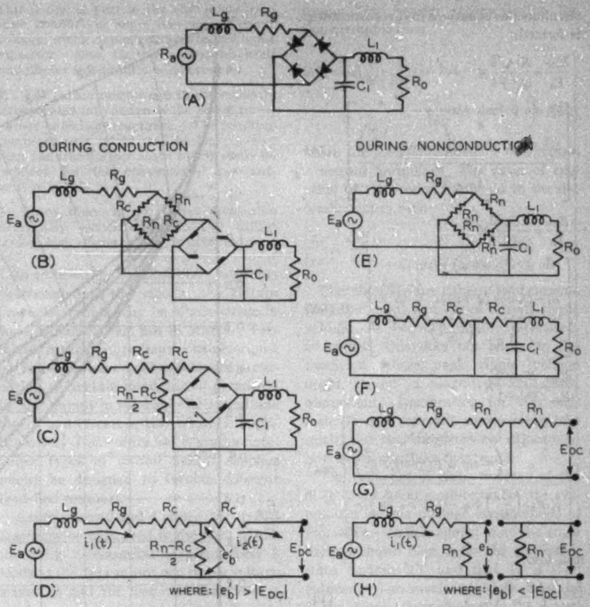


Fig. 5. Filtered inductive-ballast power supply

A—Circuit schematic
B, C, D—Successive steps in the derivation of its equivalent circuit during conduction
E, F, G, H—Successive steps in the derivation of its equivalent circuit during nonconduction

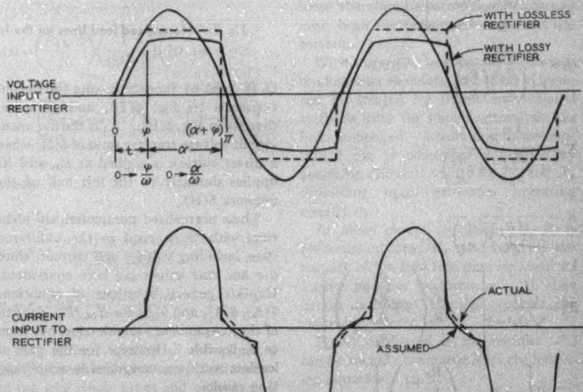


Fig. 6. Typical wave forms for the filtered inductive-ballast power supply showing the effect of rectifier losses and assumption of current discontinuity differences have been accentuated

simultaneous equations in three unknowns is formed:

$$\left\{\begin{aligned}
&\frac{E_{DC}}{E_a}=\frac{K_1\sqrt{2}}{Q_1^2+1}\times\left[\frac{Q_1\sin\psi+\cos\psi)\sin\alpha-(Q_1\cos\psi-\sin\psi)\left(\cos\alpha-\epsilon^{-\frac{\alpha}{Q_1}}\right)}{1-\epsilon^{-\frac{\alpha}{Q_1}}}\right] && \text{(7A)}\\
&\frac{E_{DC}}{E_a}=\frac{K_2\sqrt{2}}{Q_2^2+1}\times\left[\frac{(\sin\psi-Q_2\cos\psi)+(\sin[\psi+\alpha]-Q_2\cos[\psi+\alpha])\epsilon^{-\left(\frac{\pi-\alpha}{Q_2}\right)}}{1+\epsilon^{-\left(\frac{\pi-\alpha}{Q_2}\right)}}\right] && \text{(8A)}\\
&\frac{E_{dc}}{E_a}=\frac{K_1\sqrt{2}}{Q_1^2+1}\times\left[\frac{(Q_1\sin\psi+\cos\psi)(1-\cos\alpha)+(Q_1\cos\psi-\sin\psi)\times\left(-\sin\alpha+Q_1-Q_1\epsilon^{-\frac{\alpha}{Q_1}}\right)}{K_1K_3\pi\left(\frac{Z_g}{R_o}\right)+\alpha-Q_1\left(1-\epsilon^{-\frac{\alpha}{Q_1}}\right)+K_1K_4(\pi-\alpha)}\right] && \text{(9A)}
\end{aligned}\right.$$

and also:

$$\frac{I_{DC}}{I_s}=\left(\frac{E_{DC}}{E_a}\right)\left(\frac{Z_g}{R_o}\right) \tag{10}$$

where

$$Q_1=\frac{Q_g}{1+\frac{R_c}{R_g}\left[\frac{1-\frac{R_c}{R_n}}{1+\frac{R_c}{R_n}}+1\right]}$$

$$Q_2=\frac{Q_g}{1+\left(\frac{R_c}{R_g}\right)\left(\frac{R_n}{R_c}\right)}$$

$$K_1=\frac{1-\frac{R_c}{R_n}}{1+\frac{R_c}{R_n}\left(1+2\frac{R_g}{R_c}\right)}$$

$$K_2=\frac{1}{1=\left(\frac{R_g}{R_c}\right)\left(\frac{R_c}{R_n}\right)}$$

$$K_3=\frac{1}{\sqrt{Q_g^2+1}}\left[1+2\frac{R_c}{R_g}\left(1+\frac{\frac{R_g}{R_c}+1}{\frac{R_n}{R_c}-1}\right)\right]$$

$$K_4=\frac{R_c}{R_n}\left[2+\frac{R_g}{R_c}+2\left(\frac{\frac{R_g}{R_c}+1}{\frac{R_n}{R_c}-1}\right)\right].$$

Most of these derived parameters have an easily identified physical significance.

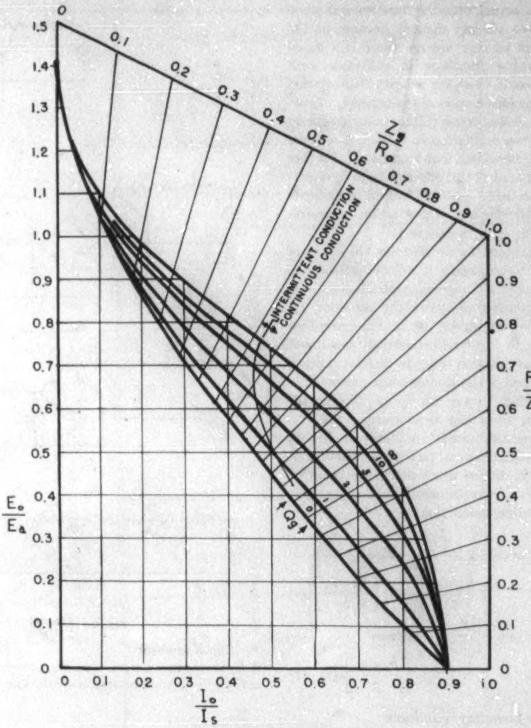


Fig. 7. Normalized load lines for the lossless filtered inductive-ballast power supply

Q_1 is given by short-circuiting the output terminals in Fig. 5(D), and Q_2 arises directly in Fig. 5(H). K_1 is the d-c open-circuit voltage transfer ratio of 5(D) when a direct voltage is applied at E_a, and K_2 applies similarly to the left half of the network 5(H).

These normalized parameters are identical with those used in the unfiltered case, including voltage and current, since d-c and rms values are here equivalent. Explicit general solutions of equations 7(A), 8(A), and 9(A) for E_{DC}/E_a in terms of the independent variables does not seem to be feasible. However, for the case of lossless rectifiers, considerable simplification results.

Equation 7(A) leads to

$$\frac{E_{DC}}{E_a}=\frac{\sqrt{2}}{Q_g^2+1}(Q_g\cos\psi-\sin\psi)\coth\left(\frac{\pi}{2Q_g}\right) \tag{7B}$$

for

$$\alpha=\pi;\quad 0\leq\frac{E_{DC}}{E_a}\leq\sqrt{2}\sin\psi$$

and equation 8(A) yields

$$\frac{E_{DC}}{E_a}=\sqrt{2}\sin\psi \tag{8B}$$

for

$$0\leq\alpha\leq\pi$$

It is thus seen that equation 7(B) holds for the mode of continuous conduction where α is known, and that during interrupted conduction when α is not known, the simple relationship in equation 8(B) holds. Solution for different values of Q_g is given in Fig. 7.

The presence of losses in the rectifier will result in a contraction of these curves. As in the previous case, the extent of this contraction is conveniently measured by the location of its end points.

At short circuit $E_{DC}/E_a=0$. Equation 8(A) requires that $\alpha=\pi$, and equation 7(A) that $\tan\psi=Q_1$. Multiplying through equation 9(A) by Z_g/R_e then gives

$$\left.\frac{I_{DC}}{I_s}\right|_{\frac{R_e}{Z_g}\to 0}=\frac{2\sqrt{2}}{\pi}\left(\frac{1}{K_3\sqrt{Q_1^2+1}}\right) \tag{11}$$

Open-circuit voltage is considerably more difficult to evaluate because current drains back through the rectifier during periods of nonconduction. A finite amount of charge must therefore be delivered to the filter each cycle to balance this leakage. The $\alpha=0$ case will be solved here despite the nonexistence of this condition. At open circuit $\alpha>0$, and a negative resistance termination would be required in order that the assumed condition exist.

These values are found to be

$$\left.\frac{E_{DC}}{E_a}\right|_{\alpha=0}=\left[\frac{2}{1+\cancel{2\frac{R_e}{R_n}}+\left(\frac{R_g}{R_n}\right)^2(Q_g^2+1)}\right]^{1/2} \tag{12}$$

where

$$\left.\frac{R_g}{Z_g}\right|_{\alpha=0}=-\frac{K_4}{K_3}$$

The value of E_{DC}/E_a at open circuit is slightly less than this zero conduction value. If a closer estimate is required, values for α can be assumed and a curve drawn to intersect the axis. Solution of equations 7(A), 8(A), and 9(A) for a specified set of conditions would seem to be best approached by assuming a set of values for α such as 0, 30, 90, 150, 160, and 170 degrees to give a good distribution of points in the upper region of the curve. The choice of α values for the lower region will depend upon the amount of rectifier losses. For the lossless rectifier, $\alpha=180$ degrees for a considerable range of values as is shown in Fig. 7. Therefore in this range, it is more appropriate to select ψ. For finite values of R_n, $\alpha=180$ degrees only at short circuit.

Design Considerations

It is now possible to see what factors make the filtered inductive-ballast power supply a good basic design.

1. Rectifier back voltage is minimized in this circuit because its output voltage is no higher (ignoring filter losses) than that required by the load itself. Moreover, conduction angles tend to be large so that peak current values are small.

2. Operating conditions are also exceptionally favorable for the filter components. This is due in part to the afore-mentioned large conduction angle and small output voltage, which minimize stress on the input capacitor. Also, the ballast reactor itself contributes significantly to filtering.

3. The ballast reactor can be conveniently incorporated into design of the power transformer as leakage reactance. The resulting combined unit might well be less massive than the transformer alone which would be required for the conventional constant-voltage supply.

4. Since there is no actual dissipation necessarily associated with the ballast action, high efficiency may be achieved.

In addition to these definitely favorable characteristics, the circuit has certain properties which may be of advantage in some situations but not in others. Terminal transient impedance considerations, for instance, have been mentioned previously. A certain flexibility is permitted by this circuit in the choice of a load line shape, and of course this circuit is applicable only if one of its loci is appropriate. Other types of virtual ballast supplies might be designed to furnish different load line contours.

Assumptions 3 and 4 under "Analysis" may now be examined in the light of the foregoing. Relaxation of assumption 3 leads to the appearance of a ripple voltage across C_1 and the load will then have a ripple component which it may be important to evaluate. Also, however, it will be seen that this ripple influences the d-c load line.

Considering first the magnitude of ripple input to the filter, solution of equation 7 for current input to the filter, assuming ripple free operation is

$$i_2(t)=\left(\frac{E_a\sqrt{2}}{R_2}\right)\left(\frac{1}{Q_1^2+1}\right)\Bigg([Q_1\sin\psi+\cos\psi]\sin\omega t-[Q_1\cos\psi-\sin\psi]\times\left[\cos\omega t-\epsilon^{-\frac{\omega t}{Q_1}}\right]\Bigg)-\frac{E_{DC}}{R_1}\left[1-\epsilon^{-\frac{\omega t}{Q_1}}\right] \tag{13}$$

where

$$\psi\leq\omega t\leq(\psi+\alpha)$$

$$R_1=R_e+\frac{(R_g+R_e)(R_n-R_c)}{R_n+R_c+2R_g}$$

$$R_2=R_g+2R_e+\frac{2R_c(R_e+R_g)}{R_n-R_c}$$

This current applied to the physical filter network will give an approximation to the true ripple input and output voltage of the filter. For most purposes, however, a useful solution can be obtained by very much simpler means. $i_2(t)$ can be fairly well represented as a full-wave rectified sinusoid when α is large. The Fourier representation for this current is then

$$i_2(t)=I_m\left(\frac{2}{\pi}-\frac{4}{3\pi}\cos 2\omega t-\frac{4}{15\pi}\cos 4\omega t\ldots.\right) \tag{14}$$

where the predominant ripple component is second harmonic. The ratio of rms value of the second harmonic to the d-c component is then

$$\frac{I_2}{I_{DC}}=\frac{\sqrt{2}}{3} \tag{15}$$

Therefore, for any value of load current, the filter input and hence output ripple voltage can be very rapidly estimated. Of course, this does not apply to the condition where peak ripple voltage might exceed in magnitude the direct component. Justification for this estimate is not so valid when α becomes small, but the approach still appears to give useful results in that region.

The influence of ripple voltage at the filter input upon equilibrium of the system can be seen through consideration of Fig. 4(F). It is assumed that d-c equilibrium shown here is for the ripple-free state (infinite C_1) and that C_1 is then reduced. The resulting ripple must be of the general form shown by the dotted line. It will be remembered that current i_1 is proportional to the integral of the shaded area, and the presence of ripple increases this area. The average value of i_2 must therefore increase and hence a higher value for I_{DC}. A new equilibrium must therefore be established which will raise both the terminal voltage and the current.

With respect to assumption 4 (dissipationless filter elements), the effect of losses can be treated by simple conventional methods after the foregoing analysis has been completed. Errors due to assumption 2 can be somewhat minimized by assuming different values for R_e and R_n depending upon estimated operating conditions.

At short circuit (equation 11), R_e is predominant over R_n in influencing departure of the load line from its idealized lossless rectifier performance. At open circuit (equation 12), R_n becomes the controlling factor.

Power dissipated in the rectifier (P_r) can be roughly estimated from the following equation

$$P_r\approx 2\frac{E_a^2}{R_n}+2.4I_s^2R_e \tag{16}$$

Power dissipated in each arm will be one fourth of this value. Over-all efficiency of the system (eff) can be approximated as follows:

$$\text{eff} \approx \frac{1}{1+2\dfrac{R_o}{R_n}+1.2\left(\dfrac{R_g+2R_c}{R_o}\right)} \qquad (17)$$

When $R_n >> R_o$ and $(R_g+2R_c) << R_o$, then efficiency will approach 100 per cent.

Power-line current is nearly sinusoidal (equation 13) under conditions of continuous rectifier conduction, and so an approximate power factor expression can be derived for this range:

$$\text{Pf} \approx \cos\psi \qquad (18)$$

where, for the lossless case, ψ ranges from a value of $(\tan^{-1} Q_g)$ at short circuit to the value

$$\psi = \tan^{-1}\left[\frac{Q_g}{1+(Q_g^2+1)\tanh\left(\dfrac{\pi}{2Q_g}\right)}\right]$$

at the critical value for continuous conduction.

Conclusions

Several circuits can perform the dual function of a ballast resistor and a conventional power supply. One of these circuits, the filtered inductive-ballast power supply, has been analyzed in sufficient detail for most design purposes. A unit of this type would be appreciably less expensive to build and to operate than its conventional equivalent.

References

1. FUNDAMENTALS OF DISCHARGE TUBE CIRCUITS (book), V. J. Francis. *Methuen's Monographs*, John Wiley & Sons, Inc., New York, N. Y., 1948, chaps. 1, 2, 6 and 7.

2. MINIATURE SELENIUM RECTIFIER HANDBOOK (pamphlet). Federal Telephone and Radio Corporation, Clifton, N. J., 1950.

3. SELENIUM RECTIFIER STACKS (pamphlet). *GET-2350A*, General Electric Company, Schenectady, N. Y., 1955.

4. COMMUNICATION NETWORKS, VOL. II (book), E. A. Guillemin. John Wiley & Sons, Inc., New York, N. Y., 1935, p. 160.

5. ANALYSIS OF RECTIFIER OPERATION, O. H. Shade. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 31, no. 6, pp. 341–61.

The Control of Air Ionization and Its Biological Effects

W. WESLEY HICKS
NONMEMBER AIEE

JOHN C. BECKETT
MEMBER AIEE

Synopsis: After eight years of research, it is clear that ionized atmospheres have demonstrable effects on human beings as well as on animals. An explanation of these effects has been sought more recently at the cellular level where results to date show significant changes. Investigation is proceeding on all levels for the purpose of clarifying results already observed and establishing dosage and other techniques which will make artificial ionization a valuable resource for human well-being.

IONS in the atmosphere have been a subject of bioclimatical investigation for many years. There is widespread belief that this is an important factor in the relationship of climate to disease. Investigation has been slow because the connection between climate and disease is complex and not one of obvious cause and effect. Acceptable proof that atmospheric ions have a biological effect has required numerous experiments under careful control of climatic factors.

Initially, these experiments were concerned primarily with the condition of ionization of air in living spaces; however, the growing scientific interest in high-altitude flying has focused attention on the need for knowledge of ionizing forces, submicroscopic charged particles, and the physiological effects of ions on man and air-borne organisms.

Ionizing Forces

The principal ionizing forces near the earth's surface are well known.[1] Of these, radioactive elements in the soil and cosmic rays are the most important. Ultraviolet radiation, X-ray, high-voltage discharge, and thermionic sources influence room ionization and therefore cannot be ignored.

Space charge, created by the excess ions of one polarity over those of the other, is an environmental factor.[2] This space charge is normally slightly positive and is attributed to the lower mobility of light positive ions in relation to light negative ions. This difference in mobility between light positive and negative ions assumes considerable importance in ventilation and air conditioning. Negative ions are removed by contact with conducting surfaces at a faster rate than are positive ions, with the result that there is selective loss of negative ions, leaving the air with an excess of positive charges.

Ion Generators

Methods of producing and controlling ionization of the air have been reported.[3] Of these, ionization from radioisotopes has been the most convenient for controlled experiments simply because all other factors could be held constant, thus reducing the possibility of technical error. For example, ionization from high voltage sources has been used, but with these devices minute amounts of ozone or nitrous oxides were also formed. Hot filaments give off thermal energy, often not desired, and they may affect particulate matter in air as the air passes near the filament.

In the experiments here reported, radioactive polonium 210 in the foil form was used as the ion source.[4] The limited range of alpha particles in air (less than 4 cm), no beta radiation, and no measurable gamma radiation make this material an ideal ion generator for experimental purposes.

The generation of unipolar ions from radioactive materials can be accomplished in several ways, with the use of an electrostatic field of low intensity for selective removal of positive or negative ions and, when desired, a current of air. For most experiments simple ion separation by means of an electrostatic field alone is preferred since a movement of air might affect test results. Fig. 1 illustrates the type of tube used in the experiments reported here. The voltage employed for ion separation varied from 20 to 300 volts, depending upon the nature of the experiment and the distance between electrodes.

Paper 57-73, recommended by the AIEE Electrical Techniques in Medicine and Biology Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 24, 1956; made available for printing January 8, 1957.

W. WESLEY HICKS and JOHN C. BECKETT are with the Wesix Electric Heater Company, San Francisco, Calif.

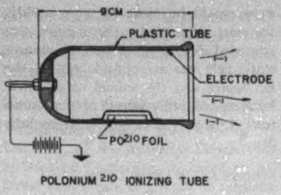


Fig. 1. Sectional view of polonium-type air-ionizing tube

Ion Measurement

There are at least two means of making ion measurements: one is to collect the ions, and the other is to measure the potential resulting from space charge. Ion measurement by a method of collecting ions and reading the current provides a direct means of determining the absolute ion density for each polarity. A single collector is limited to measuring low-mobility ions or high- and intermediate-mobility ions, since the range of potential required to measure all mobilities in one instrument is so great as to cause error.

A simple ion collector is shown schematically in Fig. 2. Parallel plates are arranged in a rectangular duct. Choice of plate size is important. Large plates will have too much capacitance for convenient operation of associated equipment. Small plates may limit the number of ions collected per second to less than the sensitivity of the current-measuring instrument, or may require a collecting voltage too high for good stability

Air is drawn past the plates with a small blower. It is important that the air velocity be kept low to avoid turbulence. A velocity of 100 cm per second is convenient for a small collector. Insulation of the collecting plates must be of a high order. Plastic insulation should be used with surface leakage path of at least 2 cm, and preferably 3. A small heater (25 watts) placed close to the pedestal insulator is used to minimize condensation when high-humidity warm air passes through the collector.

The polarizing plates should be insulated from the duct, and spaced appropriately between the collecting plates. Any convenient spacing may be used; however, a spacing great enough to minimize turbulence and prevent short-circuiting by particles of lint and dust is desirable. The spacing should not be so great that the collecting potential required on the polarizing plates will cause ionization within the device and introduce error.

Ion mobilities vary over a wide range. Light- or high-mobility negative ions move at a rate as fast as 2.2 cm/sec/volt/cm, and positive ions at about 1.6 cm/sec./volt/cm.[5] Intermediate ions move in a range of 0.1 cm/sec/volt/cm to 0.01. To measure ions of different mobilities it is necessary to vary the collecting polarizing potential. A satisfactory ion collector for most biological work will collect ions with mobilities down to 0.016 cm/second/volt/cm. In clean air high-mobility ions predominate and only near urban areas or places of air contamination are found lower-mobility ions in any great number. Incidentally, ion mobility measurement appears to be a practical means for detecting air pollution.

In addition to the ion collector, a sensitive micro-micro ammeter is necessary and a suitable recorder is helpful in making ion density measurements. Several types of commercial micro-micro ammeters are available, but it is preferable to choose one of high stability and rugged components for portable use. For this work two types have been used, a Beckman d-c Ultrohmeter, and a Beckman model V vibrating reed unit in combination with a 50-millivolt potentiometer-type recorder.

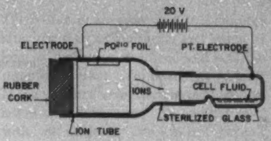


Fig. 3. Air-ionizing tube adapted for use with cell culture apparatus

Ion measurements at the earth's surface vary from hour to hour and from day to day, and it is essential to take continuous recordings, as individual readings have little meaning. These recordings should be in duplicate: positive and negative or, better, in quadruplicate: positive, high, and intermediate; and negative, high, and intermediate, where a clear picture of ion mobilities is an essential part of the experiment. A timing device to take periodically a zero reading in case of changing leakage current is a good precaution.

The space charge ion-measuring equipment has been described.[6] It has the disadvantage of measuring only the net ion density. This is offset in part by the fact that it provides a reading independent of different ion mobilities, and measures net ion density of total ions: light, intermediate, and heavy. Great care must be exercised in its use to avoid electrostatic fields from other equipment or statically charged dielectric material. For this reason it is not very satisfactory for use in biological or clinical experiments, but of real value in atmospheric studies.

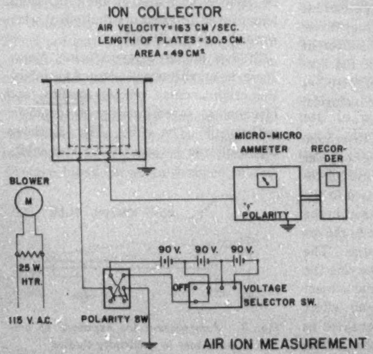


Fig. 2. Air ion measurement equipment shown schematically

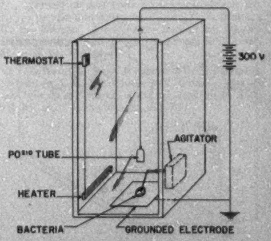


Fig. 4. Arrangement for exposure of bacteria to unipolar ions

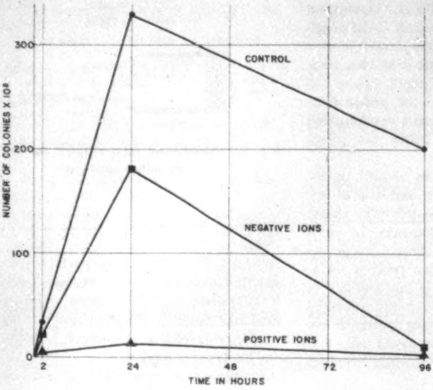


Fig. 5. Air-ion effect on growth of *Neurospora* at different times of exposure

Behavior of Submicroscopic Aerosols in Atmosphere Containing Ions

One approach to the study of climate and health is to investigate the behavior of small (submicroscopic) particles when suspended in air, particularly, particles that are known to influence respiration. Specifically, the technique has been used to follow variations in the electric charges of such particles when they are exposed to small concentrations of unipolar air ions.

The technique makes use of bacteriophage as a tracer substance. The bacteriophage, which is a small, quasi-living organism, will attack susceptible bacteria and leave visible evidence of its presence in the form of small holes in an otherwise uniform bacterial growth.

The sensitivity of this tracer technique is probably comparable to that of radioactive methods. Decontamination is simplified, since the bacteriophage has a half-life suspended in air at room temperature of about 2 hours. The air-ion environment is not altered as it would be by radioactive methods.

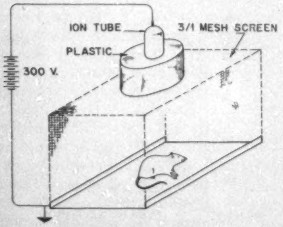


Fig. 6. Typical arrangement for exposure of small animals to unipolar ions

The investigation has used particles of a 0.2- to 0.5-micron average radius and the results indicate th[illegible] particle discharging and charging by ion densities of 1,000 to 10,000 ions/cc (per cubic centimeter) is slow. Small effects upon electrostatic precipitation rates appear to be perceptible, however, and some evidence has been obtained which suggests that exposure to low ion densities may reduce the average charge of the aerosol, but that higher ion densities may cause actual charging of the particles.[7]

Biological Investigations

With the use of radioactive ion generators for the production of unipolar ions, biological investigations were initiated on a broad front in an effort to uncover appropriate research fields. As might be expected, some of the earliest tests were inconclusive, some suggestive, and others negative. When competent assistance became available in the specialties holding promise, specific studies were pursued to determine the biological influence of air ions.[8]

Many tests were designed to determine the direct biological influence of unipolar ions. The desired ions were directed toward the organism or animal used in the experiment. Ion densities produced in the air adjacent to the subject were in the order of 20,000 to 250,000 ions/cc. The ion density was controlled by varying the field or the distance between the subject and the source. Generally, an effort was made to use as high densities as practical on the assumption that differences would be observed more readily. In most experiments the subject was grounded electrically. Thus, an electrostatic field was established between the ion source and the animal or organism being tested; this helped to direct ions toward the subject to further ensure high density.

Studies at Cellular Level

Preliminary studies produced evidence that the biological effect of ions is a fundamental one occurring at the cellular level. Whether this might be due to electric charge of the cell membrane or the change in conductivity of cell structure is not known. Continued experiments using ion generators adapted for tissue culture procedures (Fig. 3), have shown that an excess of ions of one polarity is essential for physiological effects.[9] Determinations were run in sets of 36 replicates, 9 used to determine initial counts and uniformity of planting, and 9 placed in each of three incubators at 37.5 C for 72 hours. Atmospheric ion conditions were normal in one incubator, elevated negative ion concentration in the second, and elevated positive ion concentration in the third. Statistically significant differences in rate of proliferation occurred: reproduction was accelerated under conditions of negative ionization and regarded with positive ionization.

Experiments have been conducted by A. P. Krueger and his assistant, Richard Smith, at the University of California, dealing with the effects of positive and negative ions on micro-organisms suspended in small volumes of various types of fluids; see Fig. 4. Effects were found to depend upon the nature of the test organism and the composition of the suspended fluid. Under some conditions a pronounced protective effect of ions could be demonstrated and, on the other hand, ions proved to exert a lethal effect when conditions were altered.

Studies are in progress using *Neurospora*, an organism well adapted for inhibition experiments, genetic studies, and biochemical investigations of metabolic pathways.[10] The chart, Fig. 5, shows that exposure to positive air ions inhib-

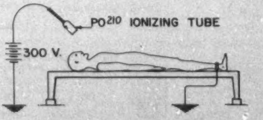


Fig. 7. Arrangement for exposure to unipolar ions in respiratory studies

ited germination of a large number of spores, and that exposure to negative air ions was also detrimental to *Neurospora* but to a lesser degree.

Growth Effects with Small Animals

Preliminary reports from Charles H. Danforth, of the Anatomy Department at Stanford University, indicated an increase in the rate of growth of chickens under negative ionization as compared with those under positive ionization or in untreated atmosphere. These findings led to investigations at the glandular level. R. J. Holloway, working at the University of San Francisco, investigated changes in cholesterol and ascorbic acid which occurred in adrenals of rats exposed for periods up to 72 days in positive and negative atmosphere. When compared to normal controls, the animals maintained in the ionized atmosphere showed changes in growth, in weight, and in cholesterol and ascorbic acid content of the adrenal gland.

Later work on the adrenal gland of rats revealed a statistically significant decline in adrenal succinoxidase activity when the animals were exposed to positively ionized air, an effect opposite to that of the stimulus of an electric shock.[11] Negatively ionized air produced no significant results. The test procedure for exposure to air ions is shown in Fig. 6.

Studies were conducted by John Allardyce at the University of British Columbia to determine the effects of negatively ionized air on the blood pressure of female Wistar albino rats. This investigator first raised their pressure 60 millimeters through injecting 0.00001 milligram of nicotine, on five successive days. The blood pressure reached a peak 16 days later and then began to recede toward normal. In the sixth week of this experiment, when pressures were still elevated, 48 of the rats were exposed to negatively ionized air for 7 days. Blood pressure of this group dropped to normal a week sooner than that of the others, and later reached a low of 20 mm below normal.

In a complementary experiment, 12 rats were exposed first to negatively ionized air and thereafter to nicotine. Average blood pressures dropped in the first 8 days from 163 to 135 mm. When nicotine was then injected, blood pressure did not rise immediately as would have been expected, but remained for 12 ensuing days at 135 mm.

Clinical Studies

Though a study of the biological effect of atmospheric ions is helpful from the scientific point of view, animal and cell studies do not in themselves indicate a medical application for this knowledge. Clinical studies designed primarily to observe effects on respiration have been conducted, with radioisotopes for an air ion source.

In a prelininary study on patients with hay fever and asthma, 27 individuals were exposed to the influence of negative ionization.[12] One third of these reported complete relief, though they had no way of knowing that the air was ionized during part of the test. A final report of this work by Dr. Piersol and Dr. Kornblueh, of the University of Pennsylvania, was made public on October 2, 1956, in Philadelphia. Results were dramatic; one third showed complete relief, and one third partial relief or improvement.

At the Good Samaritan Hospital in Los Angeles, Travis Winsor has been conducting numerous clinical tests extending over a period of 3 years. One of these studies was designed to elucidate in an objective manner the cause of nasal obstruction occurring after exposure to positive ions.

The maximum breathing capacity appears to be the most useful guide. A mask is placed over the face covering the nose. The mouth is kept closed by means of adhesive tape. The patient then breathes in and out through his nose as rapidly and as deeply as he can for one minute. The amount of air moved in this time is graphed. Exposure to positive ions, lasting approximately 5 minutes, resulted in a reduction of breathing capacity of from 35 to 25 liters per minute.

The discomfort during exposure to positive ions was consistent. In one test 16 subjects reported headaches, nasal obstruction, husky voice, dry throat, or dizziness. The same subjects reported no change during control tests. The irritation appeared in 2 to 3 minutes after the start of ionization and lasted as long as the ions were introduced into the respiratory tract. In about one quarter of the patients, symptoms lasted for some hours after ionization. In another test, with 20 subjects, two control periods were employed, one before and one after ionization. Exposure to positive ions continued for 10 minutes. Results established that symptoms persisted after ions were removed. Further tests revealed that the incidence of irritation was higher when the body was connected to ground, as shown in Fig. 7. Negative ions were also tested, and in a small percentage of subjects they also produced irritation of the upper respiratory tract, but statistically this was not significant.

References

1. Climate Control Through Ionization, T. L. Martin, Jr. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 254, no. 4, Oct. 1952, p. 269.

2. Air Ionization as an Environment Factor, J. C. Beckett. *AIEE Transactions*, vol 73, pt. II, Sept. 1954, pp. 161–65.

3. Control of Air Ion Density in Rooms, H. H. Skilling, J. C. Beckett. *Journal*, Franklin Institute, vol. 256, no. 5, Nov. 1953, pp. 423–34.

4. Production of Unipolar Air with Radium Isotopes, Thomas L. Martin, Jr. *AIEE Transactions*, vol. 72, pt. I, 1953 (Jan. 1954 section), pp. 771–76.

5. Fundamental Process of Electrical Discharge in Gases (book), Leonard B. Loeb. John Wiley & Sons, Inc., New York, N. Y., 1939, p. 5.

6. Measuring and Recording Atmospheric Electrostatic Potential, J. S. Carroll, S. B. Hammond, E. H. Stewart. *AIEE Transactions*, vol. 74, pt. I, Sept. 1955, pp. 517–20.

7. A Bacteriological Technique for the Experimental Investigation of the Ageing and Electrostatic Precipitation of Sub-Microscopic Aerosols from Ionized Atmospheres, J. A. Jamieson. *Thesis*, Stanford University, Stanford University, Calif., June, 1956.

8. Air Ion Generation, Separation, Metering and Physiological Effects, W. W. Hicks. *Journal*, Franklin Institute, vol. 261, no. 2, Feb. 1956, pp. 209–17.

9. Air Ion Concentration and the Growth of Cells in Vitro, J. L. Worden, J. R. Thompson. *The Anatomical Record*, Philadelphia, Pa., vol. 24, no. 2, Feb. 1956, p. 500.

10. Studies of Ionization Effects, Robert Fuerst. *Annual Report*, University of Texas, M. D. Anderson Hospital and Tumor Institute, Houston, Tex., 1955, p. 60.

11. Effect of Air Ions on Succinoxidase Activity of the Rat Adrenal Gland, C. B. Nielsen, H. A. Harper. *Proceedings*, Society of Experimental Biology and Medicine, Utica, N. Y., vol. 86, 1954, pp. 753–56.

12. Artificial Air Ionization in Physical Medicine, I. H. Kornblueh, J. E. Griffin. *American Journal of Physical Medicine*, Baltimore, Md., vol. 34, no. 6, Dec. 1955, pp. 618–31.

Speed Control of a D-C Motor Using a Magnetic Amplifier

W. LEONHARD
ASSOCIATE MEMBER AIEE

Synopsis: Magnetic amplifiers are, similar to thyratrons, suitable for adjustable-voltage drives of d-c motors. While the operation with discontinuous armature current affects the motor characteristics undesirably, it offers in the lower power range an important economical advantage. After a discussion of the open-loop characteristics of the drive, the design of a medium-accuracy (5-per-cent) speed-control system for the fractional- and integer-horsepower range is explained where utmost simplicity and the use of inexpensive components are considered important design factors. Performance characteristics are given.

THE no-load speed of a d-c motor is porportional to the armature voltage and inversely proportional to the magnetic flux. In order to change the speed either one or both of the quantities can be changed. The classical Ward-Leonard drive utilizes the first possibility by supplying a d-c motor, whose field is fixed, from a d-c generator, which is driven at nearly constant speed and whose excitation is controlled. The whole speed torque (n-T) plane of Fig. 1 is accessible without switches in the power circuits and with continuous transfer from one quadrant to another. The drive operates with reasonably low losses, but the initial expenses are high because three machines are necessary.

For many years, d-c motors also were successfully operated using controlled thyratrons or mercury-arc converters for supplying the adjustable direct voltage, thus replacing the prime motor and the control generator by static devices.[1–6] A system employing two converters with several anodes and a transformer with two secondaries is roughly equivalent to a Leonard drive; however, it has faster response to speed commands and disturbances and better efficiency, even though the motor losses are increased due to the ripple in the armature voltage and current. Because of the rectifying property of the converters and the induced armature voltage, the speed-torque characteristics in the low-torque region are normally not flat as in Fig. 1, but curved similarly to those of a series motor, especially if the number of supply phases is low. In order to prevent this, either a sizable choke in the armature circuit or a closed-loop speed control is required. Thyratron drives are at a disadvantage in applications where ruggedness and long life is of special importance because of the converters and the electronic control apparatus.

More recently, the problem of economically controlling the speed of small- and medium-sized d-c motors (<20 kw) has also been solved using magnetic amplifiers for converting a constant a-c supply voltage into an adjustable d-c armature voltage.[7–10] This provides ruggedness with simplicity and no warm-up time. The speed of response is usually slower than with thyratrons and the range of operation is normally restricted to quadrant I. Reversing the rotation of the motor (quadrant III) by supplying reversed armature voltage from a second amplifier is only feasible with very small motors (<100 watts) because of the poor efficiency of the circuit necessary. Inverter operation (quadrants II and IV) is impossible because it is incompatible with the operation of a magnetic amplifier. The main reason is that an amplifier element (inductor in series with rectifier) cannot block a direct voltage in the forward direction as would be required for inverter action. The exclusion of quadrant II is a more serious disadvantage than it appears at first. Even though regeneration may not be necessary in steady state because of the particular load, it is very desirable to have regeneration during transients, for example, when the speed of the motor should be quickly reduced.

The effect of the induced armature voltage and the unidirectional current supply on the motor characteristics is similar to that observed with controlled thyratrons. It is different in that magnetic amplifiers, in contrast to thyratrons, are usually connected in circuits which allow the current through the inductive load (armature of the motor) to "free-wheel" during certain parts of the cycle without affecting the amplifier cores, since lack of such provision can cause instability.[11] A magnetic amplifier in center-tap connnection with a shunting rectifier across the load would serve this purpose, but bridge circuits which, if the voltages are suitable, do not require a power transformer are more commonly used.

This paper considers the two single-phase amplifier circuits shown in Fig. 2. When S is closed, a conventional doubler circuit connected in series with a rectifier bridge results. The free-wheeling current in the armature circuit is completely decoupled from the amplifier if the rectifiers are equal. If S is open, the circuit appears as a full-wave bridge circuit. This requires two rectifiers less (*1* and *2*) and gives shorter response time in the "off" direction. However, there is a counteraction of the free-wheeling current to the amplifier due to the forward drop of rectifiers *3* and *4*. If this is neglected, both circuits are equivalent in steady state.

Simplified Analysis of the Open-Loop Characteristics

SPEED-TORQUE CHARACTERISTICS

First a brief discussion will be presented on how the speed-torque characteristics of the motor are modified when the armature is supplied from the unidirectional amplifier output rather than from a smooth direct voltage. The two different modes of operation, with continuous and discontinuous armature current, which are described in the literature[1,3,4] for thyratron circuits, are observed here also. The analysis is simplified by the following assumptions:

1. The amplifier cores have rectangular magnetization characteristics with zero loop width, their windings have zero resistance, the rectifiers are ideal ($R=0$ or ∞), the mechanical time constant of the drive is much larger than the a-c period, and the excitation flux ϕ of the motor is constant.
2. The firing angle φ of the amplifier depends only on the control current I_c and is independent of the load.

The simplifications made, especially item 2, are such that only qualitative results can be expected. With these assumptions, the circuit and the supplied voltage can be redrawn as seen in Fig. 3. The missing parts of the sinusoid are ab-

---

Paper 57-93, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 24, 1956; made available for printing November 21, 1956.

W. LEONHARD is with Westinghouse Electric Corporation, East Pittsburgh, Pa.

The author wishes to express his gratitude to R. O. Decker, Westinghouse Electric Corporation, who contributed valuable ideas and furthered the work through his encouragement.

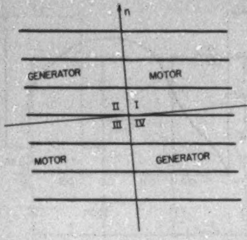


Fig. 1. Characteristics of Ward-Leonard drive

sorbed by the reactors and the remainder v_{ac}' is applied to the a-c terminals of the rectifiers. L_a is the inductance of the armature circuit; r_a is its total resistance; E is the induced electromotive force, a smooth direct voltage; i_a is the instantaneous and I_a the average armature current, while V_{ac}' is the average value of the voltage v_{ac}'. Since ϕ is constant, $E=k_n n$ and $I_a=k_t T$, where k_n and k_t are constants of the motor.

Discontinuous Armature Current

For $L_a=0$, the current in the armature flows only during that part of the period where $|v_{ac}'|>E$, i.e., there are two separate current pulses per cycle. I_a depends on φ, r_a, and $E/\hat{v}_{ac}$. The speed-torque characteristics are not as straight as those of a d-c shunt motor because in the ideal no-load condition the induced voltage E equals the peak voltage $v_{ac}'{}_{\max}$ and blocks the rectifiers completely, while in standstill the armature current I_a is determined by the average voltage V_{ac}'. The solid curves of Fig. 4 show the characteristics for several firing angles φ. The end points of each characteristic are determined by:

No load: $I_a=0;\ E=v_{ac}'{}_{\max}$

Standstill: $I_a=\dfrac{V_{ac}'}{r_a};\ E=0$

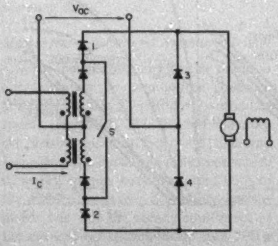


Fig. 2. D-C motor supplied from magnetic amplifier

The speed and torque in Fig. 4 are referred to the no-load speed $n_{\max}$ and the (ideal) stalled torque $T_{\max}$ of the motor when supplied with $V_{dc}=V_{ac}'$ $(\varphi=0)$ which is its rated voltage. The rated torque $T_0/T_{\max}$ of a particular motor is generally between 0.05 and 0.2, depending on the motor size.

If the armature circuit contains some inductance, the pulses of the armature current will still begin at τ_1 (Fig. 3); however, they will not cease at τ_2, because of the inductance, but will continue until later, τ_3. Further increase of L_a produces free-wheeling current during the time $\pi<\tau<\tau_4$.

Continuous Armature Current

Assume more resistance-free inductance is added so that finally $\tau_4=\tau_1+\pi$. The current then begins to flow continuously. For this case, the determination of the steady-state characteristics becomes very simple. The equation

$$E+i_a r_a+L_a di_a/dt=|v_{ac}'|$$

of the armature circuit is now valid throughout the a-c period. Averaging it over one period and realizing that the direct voltage across L_a is zero, results in the speed-torque characteristics:

$$n/n'_{\max}=1-T/T'_{\max}$$

Here $n'_{\max}$ and $T'_{\max}$ are the no-load speed and the stalled torque of the motor when supplied with a reduced direct voltage $V_{dc}'=V_{ac}'(\varphi)$. Continuous conduction produces, therefore, the original linear shunt motor characteristics. These curves are plotted as dashed lines in Fig. 4. The parameter $\omega T_a=\omega L_a/r_a$ which is required for continuous conduction depends on the speed and the torque or the firing angle. It becomes infinite when the torque approaches zero.

Fig. 5 shows a family of calculated speed-torque characteristics for $\varphi=90$ degrees and different armature time constants. It is seen that with increasing ωT_a the range of continuous current extends to lower torque. In Fig. 6, the characteristics are plotted for various φ and $\omega T_a=4$, which is a reasonable ratio of the internal armature reactance at 60 cycles to the armature resistance for a small motor. Finally, Fig. 7 shows the armature constant ωT_a, necessary to produce continuous conduction for various speeds at $T/T_{\max}=0.02$, which is, for example, 20 per cent of rated torque for a motor with 10% (per cent) rated speed drop in normal smooth d-c operation. The maximum armature time constant is required at medium speed: $\omega T_{a\ \max}\approx 33$. At 60 cps (cycles per second), this can be

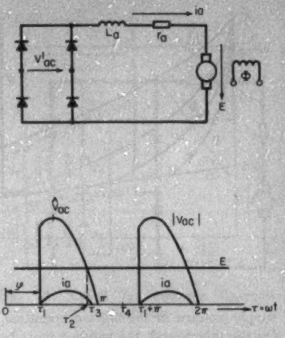


Fig. 3. Simplified armature circuit

approached only by using a filter reactor in the d-c armature circuit. The insertion of an armature reactor seems advisable indeed because it would not only produce the desired motor characteristics but it also would reduce the ripple current and so the motor losses.

An approximate determination of the physical size of the external reactor leads to interesting conclusions, however. With continuous conduction at low load, the alternating voltage appearing across the total reactance of the armature circuit depends mainly on φ. The maximum voltage is found to correspond to approximately 30% of the a-c supply voltage. Most of this voltage appears across the filter reactor. Because of the unidirectional armature current, the reactor core is magnetized continuously in one direction unless a separate bias magnetization, using for example the field current of the motor, is applied. If this bias is not provided, the reactor must be designed so that it can absorb at least 60% of the

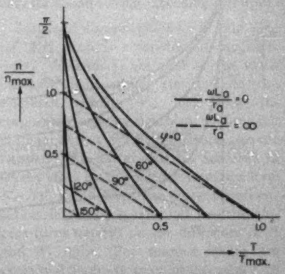


Fig. 4. Speed-torque characteristics of rectifier-supplied d-c motor

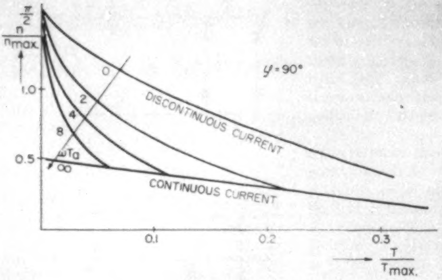


Fig. 5 (left). Speed-torque characteristics for different armature time constant ωT_a

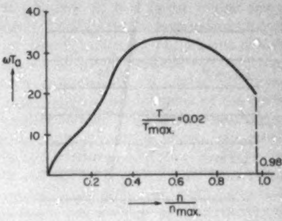


Fig. 7 (right). Armature time constant required for continuous conduction at $T/T_{max}=0.02$

supply voltage in normal a-c operation. On the other hand, the winding must be capable of handling a current $I_{rms} > I_{a\ max}$ with low losses. The reactive power rating of the reactor at 60 cps is, therefore, roughly $N_R > 0.6\ N_{motor}$. This kilovolt-ampere rating is approximately 20% higher than that of one of the main reactors of the magnetic power amplifier. However, the filter reactor is most economically designed since it saturates at $T > 0.02\ T_{max}$ during parts of the cycle. The length of these saturated periods increases with the load. This means that at higher load this reactor becomes more and more ineffective as a filtering element. The current is always continuous for $T > 0.02\ T_{max}$, but the peak current is essentially the same as without the reactor. Therefore, it cannot be expected to have much influence on the motor losses at high load or on the commutator life. If saturation is to be prevented at high load in order to maintain the filtering effect, the size of the reactor would become quite unacceptable indeed.

There are possibilities for a better solution:

1. A multiphase supply or increased supply frequency will reduce the size of the armature inductor.
2. Using a closed feedback loop, the range of discontinuous armature current can be utilized and the curved load characteristics can be straightened. The absence of a large armature reactance is then advantageous for stability.

A multiphase supply is an appropriate solution for motors above about 5 kw. In this power range symmetry of the line currents becomes increasingly desirable; it is also expected that there is a cost advantage for the multiphase amplifier.[7] The frequency usually cannot be increased. Therefore, for the power range considered in this study, it appeared advisable to choose the second approach.

CONTROL CHARACTERISTICS

Fig. 6 shows the speed-torque characteristics $n(T)$ of the motor. In order to recognize how the dependency of the speed on the load can be reduced by a closed-loop system, the other part of the function $n(T, I_c)$, i.e., the control characteristics $n(I_c)$ must be known. These characteristics also depend, of course, on whether the motor operates with continuous or discontinuous armature current. The curves, shown in Fig. 8, are constructed from those in Fig. 6, using the additional assumption that the amplifier with resistive load and proper bias has a linear transfer function $V_{out}=G_aI_c$. This links φ and I_c. The curves for $\omega T_a=\infty$, Fig. 8(A), are linear and regularly displaced due to the voltage drop across the armature circuit resistance. The control characteristics obtained for $\omega T_a=4$, Fig. 8(B), are curved and far apart for low load, but approach those in Fig. 8(A) for increasing load. The characteristic for $T/T_{max}=0.15$ lies completely in the region of continuous current and is, therefore, identical with that of Fig. 8(A).

The utilization of discontinuous currents requires, for the same steady-state accuracy, higher gain in the speed-control feedback loop than when operating with continuous armature current. This can be seen qualitatively from Fig. 8. Assume the feedback loop has a certain steady-state gain:

$$G_f=-\frac{dI_c/I_{cmax}}{dn/n_{max}}$$

The speed change of the motor under load

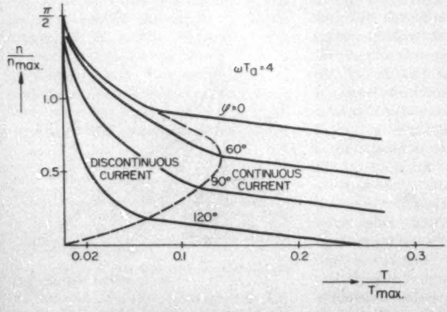


Fig. 6. Speed-torque characteristics for $\omega T_a=4$ and different firing angle φ

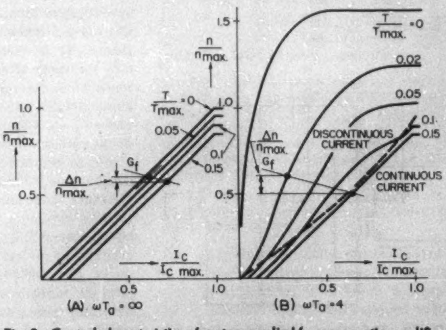


Fig. 8. Control characteristics of motor supplied from magnetic amplifier

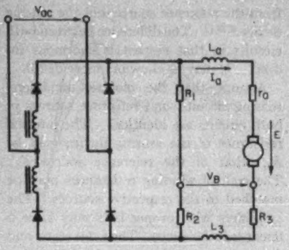


Fig. 9. Bridge-sensing circuit for induced armature voltage

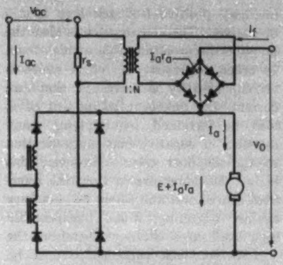


Fig. 10. Circuit for $I_a r_a$-compensation using transformer

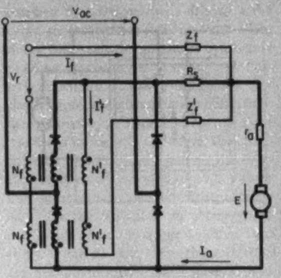


Fig. 11. Circuit for magnetic $I_a r_a$-compensation

can then be graphically constructed. The speed may be adjusted in both cases for no load ($T/T_{max}=0.02$) to $n/n_{max}=0.60$; then the load may be increased to $T/T_{max}=0.15$. It is seen that the speed variation $\Delta n/n_{max}$ is much larger in Fig. 8(B). Therefore, in order to produce the same speed change with discontinuous current, a higher feedback gain G_f must be provided.

Feedback Loop

SPEED MEASUREMENT

Feedback control requires that the speed be measured and compared with a reference quantity in order to detect deviations. In this case, no highly accurate measurement is required (2–3%); however, the sensing device must provide a sufficiently strong signal to get good speed regulation even when using a single-stage low-gain amplifier and when operating the motor with discontinuous armature current. Several possible sensing devices are discussed in the following.

Tachometer

A tachometer generator could be used for accurate speed measurement, but the costs usually prohibit its use in a medium-accuracy drive.

Measurement of Induced Armature Voltage

Since the exciting flux of the motor is approximately constant, the induced armature voltage E is proportional to the speed. However, the voltage between the brushes contains, besides E, the drop $I_a r_a$ along the armature resistance which is, at full torque, between 5 and 20% of the induced voltage, depending on the motor size. If the steady-state error of the closed-loop system must stay below this value, compensation for $I_a r_a$ is required.

Bridge Circuit. R_1, R_2, R_3 in Fig. 9 are external resistors, L_3 is an external inductor. For $R_1/R_2=r_a/R_3=L_a/L_3$ the output voltage V_B depends on E only and is independent of I_a. If the armature current is discontinuous, L_a can be neglected[6] for slow transients and L_3 can therefore be omitted. The no-load output voltage of the bridge is $V_{B0}=ER_3/(r_a+R_3)$ and the maximum output power is

$$P_{B\,max}=\frac{R_3}{R_1+R_3}\;\frac{E^2}{4r_a}$$

It can be increased by increasing R_3 or by decreasing R_1+R_2. Both measures boost the power dissipated in these resistors.

This circuit, used by R. O. Decker in a previous development, is very advantageous for its simplicity and accuracy. Its application, however, is restricted to systems with high forward gain. In low-gain systems, high losses in the ballast resistors would be necessary to produce the required system accuracy.

Voltage Compensation. Another way of measuring E is the use of a compensating device which performs essentially the function of a d-c transformer, i.e., which has the armature current (or a proportional quantity) as input and an isolated direct voltage $I_a r_a$ as output. This voltage can then be used to subtract the $I_a r_a$-drop from the armature voltage and so to sense E.

In Fig. 10 one simple example is shown. The voltage drop across the shunt resistor r_s in the a-c circuit is transformed, rectified, and with proper polarity introduced into the feedback loop, where the reference voltage and the sensed voltage are compared. With $N=r_a/r_s$, $V_0 \approx E$. This is true if $i_f>0$ is assured. Another possibility is seen in Fig. 16. There a low-gain magnetic amplifier which is controlled by the actual armature current I_a, supplies the compensation voltage $I_a r_a$. Because of the low gain required, the amplifier can be designed easily for very good linearity, for example, using negative feedback.

The first solution is simpler, but it causes unsatisfactory operation at low speed and at high load because of the free-wheeling current in the armature circuit which disturbs the relation $I_a \approx I_{ac}$ (Fig. 10). The losses in r_s are also objectionable, even though they are much lower than those in the bridge circuit (Fig. 9). The magnetic amplifier as a compensating device operates equally well at any speed and involves only negligible losses. An important advantage is, also, that the delay of the compensating magnetic amplifier tends to improve the system stability. This is a decided difference from what must be expected when a preamplifier stage is added to the main amplifier and a low-power bridge-sensing scheme is employed.

Both the transformer and the magnetic amplifier compensating schemes can easily be designed to incorporate a crude current-limiting feature. If their output-input characteristics saturate at a predetermined overcurrent, the speed regulation of the motor will increase considerably for a load torque exceeding this limit.

Magnetic Compensation.[8] In Fig. 11, the $I_a r_a$ drop is compensated magnetically. N_f and Z_f are part of the main feedback loop which can include the reference voltage V_r. This loop incorporates negative feedback of the back voltage E and of the armature current ($I_a r_a$). This latter can be canceled in steady state by the positive feedback loop R_s, N_f', Z_f'. When the current compensation is complete, the net control ampere-turns depend on the difference of E and V_r only. For reasons which are explained later, it may be necessary to insert a filter inductance L_f into the main feedback loop. In order to prevent de-

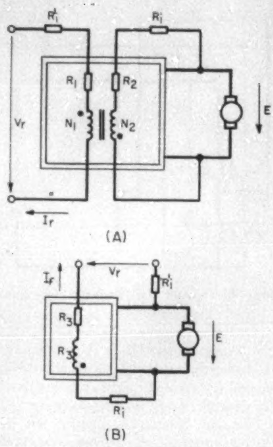


Fig. 12. Circuits for comparison of V_r and E

terioration of the stability, an inductor L_f' may also be required in the auxiliary positive feedback loop.

This method of measuring E by magnetic $I_a r_a$-compensation is advantageous because it is accurate and also allows high feedback gain since the main feedback loop is connected directly to the armature. There is, however, the following difficulty: If the reference voltage is connected in series to the armature as shown in Fig. 11, I_f is usually positive at no load, but it decreases if load is applied, and it may invert its sign if the load and the loop gain are high. If the reference voltage is supplied by rectifiers, I_f cannot change its sign, however. This results, therefore, in a partial or complete disconnection of the main negative feedback loop, while the auxiliary positive feedback loop is still operative. The consequence is that the amplifier becomes unstable at high load. In order to prevent this effect which is accompanied by an excessive armature current, the quiescent component of I_f may be increased by applying large amounts of negative bias magnetization to the amplifier cores. However, this boosts the power drawn from the reference source and still offers no certainty against triggering, if an unexpectedly high load surge occurs. Therefore, the positive feedback current I_f' must be limited. This could be accomplished, for example, by a pair of saturable reactors which are controlled by I_a and supply proportional direct current I_f' into the feedback winding N_f'. The bias magnetization on the amplifier cores may then be adjusted so that, for any speed and $I_f'=I_{f\ \max}'$, the current I_f still has a safe positive margin.

COMPARISON OF REFERENCE AND SENSED VOLTAGE

The preceding section dealt with the sensing of the speed in form of the direct voltage E. This voltage has now to be compared with the reference quantity, normally an adjustable direct voltage V_r obtained from a variable transformer and rectifiers, with the result that the control ampere-turns of the magnetic amplifier are proportional to V_r-E. There are several ways to perform this superposition. Figs. 12(A) and (B) show two of them. In Fig. 12(A), both voltages produce ampere-turns which are superposed on the amplifier cores; in Fig. 12(B), both voltages are superposed immediately. It is known that the voltage comparison, Fig. 12(B) is usually preferable since it requires less power from the reference source and the sensing device.[9,10,12] The difference between both circuits in that respect is enormous indeed. It may be shown in more detail.

Assume that the motors, amplifiers, sensing circuits, and reference sources in both circuits are identical. The internal resistance of the sensing device may be R_i; that of the reference source R_i'. The control winding resistances may be matched to the respective sources. The amplifiers with proper bias may have a linear transfer curve. Then, the static and dynamic performances of both systems can be equalized completely by proper distribution of the available control winding space into the reference winding N_1 and the feedback winding N_2. However, there is a large difference in control power. It is found that for a certain speed the powers drawn from the equal reference sources have the ratio:

$$P_{ra}/P_{rb} \approx (1+R_i/R_i')G_L$$

$G_L>>1$ is the equal loop gain of both systems. The ratio between the total powers dissipated in the control windings is

$$P_{wa}/P_{wb} \approx \left(2+\frac{R_i}{R_i'}+\frac{R_i'}{R_i}\right)G_L^2$$

This is so because the major part of the reference power in the first case is regenerated in the armature of the motor, while in the second case both contributions, the power drawn from the reference source and that from the sensing element, are dissipated. With $R_i=R_i'$ and $G_L=30$, one finds $P_{wa}/P_{wb}=3{,}600$. It is obvious indeed that the second circuit is preferable in all these applications where the economy of the control circuitry is important, i.e., in low-gain systems. The second solution also has disadvantages:

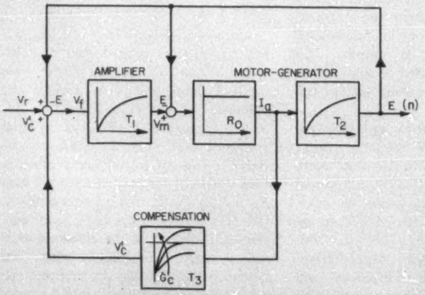


Fig. 13. Block diagram for stability calculation

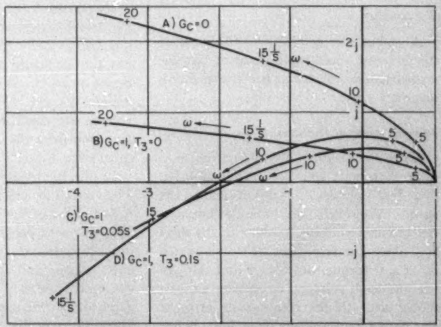


Fig. 14. Characteristic polynomials P(jω)/P(o) for different compensation

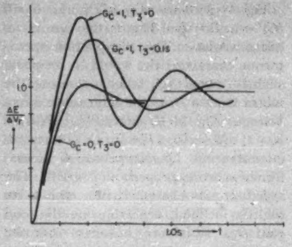


Fig. 15. Calculated speed transients following a step change in reference voltage

1. The reference voltage V_r must be isolated and must be adjustable through approximately the same range as the voltage E is required to be. The internal resistance of the reference source directly influences the loop gain.

2. When V_r is a rectified alternating voltage, the problem of discontinuous current can appear in the control loop in the form of discontinuous sensing. This has the effect of introducing a large variable resistance into the control loop which decreases the loop gain drastically. This can be prevented by inserting an inductance L_f of small volt-ampere rating in the feedback loop. However, an ample amount of negative bias on the amplifier cores increases the quiescent component in I_f and so alleviates the need for an inductor.

RESULT

The discussion of the different methods for speed sensing and feedback suggests that for a low-gain system it is best to use only one control winding on the amplifier, and to perform the $I_a r_a$-compensation and the superposition of the sensed voltage with the reference voltage by voltage addition in one circuit.

Further Design

Besides the variable load, which is the most obvious disturbance, there are some more influences, which must be expected to cause considerable deviations of the speed from a set value.

1. Temperature variations in the field winding of the motor during operation, causing changes in field current (20%) and magnetic excitation flux.

2. Supply voltage fluctuations (±10%) affecting the reference voltage and the field current of the motor. Both will usually be supplied from the same a-c line.

The armature reaction on the magnetic flux can be canceled by a compound winding. All other influences, as fluctuating bias current, changing amplifier supply voltage, and frequency variations, have second-order effects.

The excitation flux of the motor could be stabilized by a negative temperature coefficient ballast resistor. Where this is undesirable, motors with highly saturated field poles can be used. The latter possibility will not fully rule out speed drift as the motor heats up. However, it permits very simple provisions against influence 2, which is normally the more undesirable one. If the line voltage fluctuates, the exciting flux of the motor follows a function $\phi(V_{\text{line}})$, which is approximately the saturation characteristic of the motor field. Assume the reference voltage V_r is stabilized, for example, by a well-known combination of a linear inductor and a saturable transformer, such that for any speed setting the resulting characteristic $V_r(V_{\text{line}})/V_{r0}$ has in the interesting region approximately the same shape as $\phi(V_{\text{line}})/\phi_0$, then good speed constancy with fluctuating line voltage may be expected. The matching of the two characteristics is indeed feasible since their nonlinearity is produced by the same effect, saturation of iron. The saturable transformer is frequency-sensitive, but in normal applications, where 60-cps line voltage is used, this is not objectionable.

The negative bias magnetization needed for the magnetic power amplifier can be provided simply by connecting the bias windings in series with the field winding of the motor. This eliminates special rectifiers and power loss in ballast resistors. An increase of the time delay of the amplifier is avoided because the impedance of the field winding is usually large enough to prevent an appreciable circulating current during transients.

Approximate Calculation of the Closed-Loop System

After it is established which feedback scheme is to be employed, a crude calculation of the system stability will be of value. For this purpose, it may be assumed that the motor is driving a d-c shunt generator with resistive load. The motor with control equipment is a highly nonlinear carrier-type system. An "exact" calculation was cumbersome for steady state and stringent simplifications were necessary. This is even more the case for transients. Therefore, the calculation is restricted as usual to the comparatively slowly varying average values. The steady-state characteristics (calculated for the simplified model or measured) can then be applied. It is seen (Figs. 6 and 8) that these characteristics are also nonlinear. The common practice is to linearize the system around a chosen operating point and to calculate the dynamic behavior for small deviations. The disturbance will be assumed to be a small change of the reference voltage V_r, while the speed n (or E) is considered the output of the system. The calculation here will be restricted to the question of how stability and transients are affected by the $I_a r_a$-compensating device. The particular choice of the disturbance is, therefore, insignificant. A change in load would lead to the same conclusions.

Fig. 13 shows the block diagram of the linearized control system. The coordinates of the operating point are n_0, E_0, V_{m0}, I_{a0}, with V_{m0} being the "internal" output voltage of the magnetic amplifier, extrapolated for no load. The magnetic power amplifier has the time constant T_1, measured for resistive load. This is based on the total resistance of the control loop. V_f is the net control voltage of the amplifier. R_0 is the armature resistor which produces the sloped open-loop speed-torque characteristic. R_0 comprises all actual resistances (rectifiers, windings, armature) of the power circuit and also contains a hypothetical resistor which stands for the effect of discontinuous armature current, discussed earlier in detail. T_2 is the effective mechanical time con-

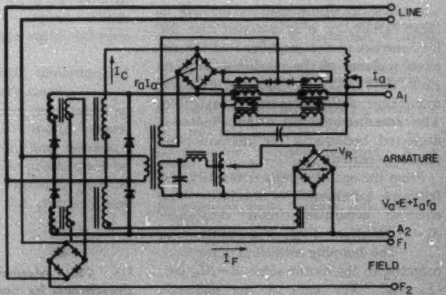


Fig. 16. Circuit diagram of d-c motor drive

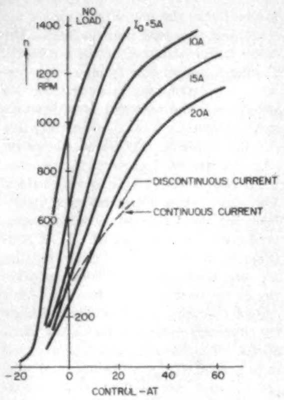


Fig. 17. Measured control characteristics

stant of the motor-generator set for the given resistance R_0 of the armature circuit and the particular load. T_2 differs from the time constant T_{20}, valid for straight d-c supply (without series resistor) and for constant torque load, by the equation

$$T_2 = T_{20}\frac{R_0/r_a}{1+R_0d}$$

d is the damping coefficient of the load, as defined in the following. For simplicity, the two feedback signals derived from the armature current, i.e., the I_ar_a drop due to sensing of the total armature voltage instead of E and the delayed output voltage V_c of the compensating device are immediately superposed with proper sign to the total current feedback signal V_c'. T_3 is the time constant of the compensating device. Its steady-state gain is $G_c = V_c/I_ar_a$. For $G_c' = (1-G_c) > 0$, the motor is undercompensated and for $G_c' < 0$, it is overcompensated.

Besides the simplifications implied in these definitions, it is assumed that the bias current of the amplifier and the excitation fluxes of the machines are constant, that $I_f \ll I_a$, and that the armature inductance of the generator is negligible. The armature inductance of the motor is neglected because of the discontinuous armature current.

From the open-loop characteristics, the following information is obtained: The effective armature circuit resistance: $R_0 = -\partial E/\partial I_a$; ($V_m = V_{m0}$ = const, E_0, I_{a0}). The damping coefficient of the load, reflected to the motor side: $d = \partial I_a/\partial E$; ($E_0$, I_{a0}). The gain of the magnetic power amplifier: $G_a = \partial E/\partial V_f$; ($I_a = I_{a0}$ = const, E_0), and the total open-loop gain: $G_L = G_a/(1+R_0d)$. The analysis of the system results in the following frequency response function of the speed versus reference voltage:

$$\frac{\Delta E}{\Delta V_r}(s) = \frac{G_L(1+sT_3)}{1+G_L(1+dr_aG_c')+s[T_1+T_2(1+r_a/R_0G_aG_c')+T_3(1+G_L(1+dr_a))]+s^2[T_1T_2+T_1T_3+T_2T_3(1+r_a/R_0G_a)]+s^3T_1T_2T_3}$$

It was mentioned before that mainly the stability problem shall be considered. For this, it is sufficient to inspect the denominator which is the characteristic polynomial $P(s)$ of the system. It is presented in the form of a plot $P(s)/P(o)$ for $s=j\omega$. With increasing ω, the curve has to encircle the origin counterclockwise, i.e., for $0 \rightarrow \omega \rightarrow +\infty$ the angle arc $P(j\omega)$ has to sum up to $m\pi/2$, to assure stability ($m=3$ is the order of the polynomial).[13] The more distant the curve encircles the origin, the larger are the real parts of the zeros and the better damped are the transients. This implies, of course, that the frequency-scale factors of the curves which are compared are not too different.

The values for the parameters are taken from an actual 3-hp (horsepower) motor drive described in the following section. For an operating point $n_0 = 800$ rpm, $I_{a0} = 10$ amperes, it was found:

$T_1 = 1.0\ s$
$T_2 = 0.38\ s$
$G_a = 47.5$
$G_L = 32$
$r_a = 0.75$ ohm
$R_0 = 5.0$ ohms
$d = 0.10$ 1/ohm

Fig. 14(A) shows $P(j\omega)/P(o)$ for $G_c=0$ ($G_c'=1$, $T_3=0$). This is the case of no compensation. The steady-state speed-torque characteristics would be dropping slightly more than when operating the motor from a constant adjustable direct voltage. Fig. 14(B) shows $P(j\omega)/P(o)$ for $G_c=1$, ($G_c'=0$), $T_3=0$, i.e., complete instantaneous I_ar_a-compensation (transformer scheme, properly adjusted). The speed-torque characteristics would be slightly inclined, depending on the loop gain G_L. It is seen, however, that the system, though stable, possesses transients with poorer damping. For $G_c \geqq 1.5$, the system would oscillate. The plots in Figs. 14(C) and 14(D) are again for $G_c=1$, but for $T_3=0.05s$ and $0.1s$. This could be obtained using a magnetic amplifier for I_ar_a-compensation. The system would have the same steady-state accuracy as before. The figures show, however, that the curves withdraw from the origin with increasing T_3, thus indicating improved damping of the transients. For the assumed load and $T_3=0.05s$, the gain G_c of the compensation device could be increased to 1.8 before instability occurs. The boundary of stability for $T_3=0.1s$ is $G_c=2.3$. It is normally, of course, of little practical use to operate more than slightly in the overcompensated range since this yields rising speed-torque characteristics. In Fig. 15, three calculated transients are shown without, with, and with delayed I_ar_a-compensation.

These results show indeed that introducing time lag into the compensation device is advantageous. For large T_3 the dynamic behavior of the system approaches that with no compensation at all.

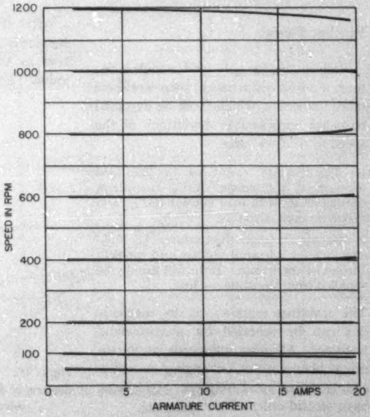


Fig. 18. Measured load characteristics for closed loop

The design of the compensating device is, therefore, quite uncritical, because for increased delay the system tends to the safe side.

Transients caused by large disturbances have a different form, at least in the beginning when the deviations are still large. This is mainly because the power transfer is unidirectional. For example, when the reference voltage is suddenly reduced to a much lower value than before, the magnetic amplifier will be partially turned off and the armature current becomes essentially zero since the induced voltage of the motor, which still runs at high speed, blocks the supplying rectifiers. Then the motor slows down according to its load and the iron and friction losses. As soon as the level of the amplifier output is reached, the armature current rises and prevents further deceleration. Then, the quasilinear transient begins. The motor "free-wheels" during the first part of the transient because there is no possibility to force the speed down by reversing the armature current (quadrant II in Fig. 1), unless a dynamic braking resistor with its disadvantages is applied. This free-wheeling effect gives the drive at low load a slow response to speed commands in the down direction.

Experimental Results

Two systems have been tested, using a 1/3-hp and 5-hp motor. The second system will be reported in detail since it lies in a power range of more practical interest.

Equipment and Circuit Used

Motor. 5-hp d-c shunt motor for 220 volts, 20 amperes, 1,750 rpm, with four poles, two commutation poles, no compound field. $\omega T_a \approx 5$, measured at 60 cps and at rated field current.

Amplifier. Consists of two transformer-type *C*-cores (Hipersil) and 5- by 5-inch selenium rectifiers. Supply voltage is 230 volts, 60 cps, d-c output current 20 amperes. The rectifiers are cooled by forced air.

Both amplifier circuits, the single-phase doubler with rectifier bridge and the bridge amplifier circuit, were tested. The bridge circuit with open feedback loop was unstable at low speed because of the inductive load and the large forward drop of the selenium rectifiers. After the loop was closed, however, the system was stable. All the further experiments were made with this circuit. The $I_a r_a$-compensation in the final circuit (Fig. 16) was performed by a small magnetic amplifier using two *U*-laminated cores and selenium rectifiers.

Measurements

The calculations, presented in the section on "Simplified Analysis of the Open-Loop Characteristics," are based on extreme simplifications which are not verified here, where selenium rectifiers and transformer cores are employed. Yet the general form of the curves is well maintained. Fig. 17 shows, for example, the measured control characteristics. The rated torque (armature current) of the actual motor corresponds to $T/T_{max} \approx 0.07$ in the normalized diagram (Figs. 6, 8). It is seen that the motor operates mainly with discontinuous armature current. The speed change between no load and full load is, therefore, very large; for example, nearly 700 rpm at zero control AT. These "open-loop" control characteristics were actually measured with closed loop because of the instability mentioned before. It is seen that the 220-volt motor cannot be driven to its top speed using a 230-volt a-c supply. This can be offset by using a suitable armature.

In Fig. 18, the closed-loop speed-torque characteristics are seen. The curves show a speed drop under load of about 40 rpm at 1,200 rpm (3–4%) and less for speeds down to 50 rpm. The speed can be smoothly controlled until a complete stop is reached. The curvature at the top speed is caused by lack of output voltage from the amplifier.

Good performance of the drive for fluctuating supply voltage is achieved by adopting a commercial variable transformer, which is used for speed setting, as a saturable transformer for the stabilization of the reference voltage (Fig. 16). Since its saturation characteristic matches approximately that of the motor field, less than 30 rpm change of the no-load speed is observed for any speed setting, when the supply voltage varies between 200 and 250 volts.

The system is slightly underdamped but safely stable for every speed.

No measurements have been taken so far to determine the motor losses, efficiency, power factor, and necessary amount of derating. The problem will be similar to that with thyratron drives with no armature reactor, but certainly less serious since the current can free-wheel and is, therefore, more smooth. Also, the "firing" of a magnetic amplifier (especially using *C*-cores) is a much less abrupt transient than the firing of a thyratron because the saturated inductance of the cores helps to reduce high harmonics in the armature current.

Conclusions

A fundamental introduction, design considerations, and laboratory test results have been given for an adjustable d-c motor drive using a standard single-phase magnetic amplifier. No tachometer generator and no inductor in the armature circuit are required. The control equipment is rugged and needs no special maintenance. The drive has a medium accuracy (<5% deviation) and allows speed control between zero and full speed. The practical power range of this scheme extends, for a single-phase supply, from fractions of a kilowatt up to probably 5–10 kw.

References

1. Zur Regelung von Gleichstrommotoren über gittergesteuerte Gleichrichter, **W. Schilling.** *Archiv für Elektrotechnik*, Berlin-Charlottenburg, Germany, 1935.

2. Electronic Speed Control of Motors, **E. F. W. Alexanderson, M. A. Edwards, C. H. Willis.** *AIEE Transactions*, vol. 57, June 1938, pp. 343–54.

3. Das Drehzahlverhalten des Gleichstrommotors bei Speisung durch gittergesteuerte Gleichrichter, **W. Schuisky.** *Elektrotechnik und Maschinenbau*, Vienna, Austria, 1939.

4. Theory of Rectifier D-C Motor Drive, **E. H. Vedder, K. P. Puchlowski.** *AIEE Transactions*, vol. 62, 1943, pp. 863–70.

5. Increased Losses in a D-C Motor When Operated from Grid-Controlled Rectifiers, **C. Richard Reiter, Charles R. Ammerman.** *Ibid.*, vol. 71, pt. II, Jan. 1952, pp. 77–82.

6. Servomechanism Characteristics of D-C Motor Driven by Controlled Rectifiers, **L. Dale Harris.** *Ibid.*, vol. 70, pt. II, 1951, pp. 1582–88.

7. Magnetic Amplifiers Invade Heavy Duty Systems, **U. Krabbe.** *Control Engineering*, New York, N. Y., 1955.

8. Magnetic-Amplifier Control of D-C Motors, **A. Kusko, J. G. Nelson.** *AIEE Transactions*, vol. 74, pt. I, July 1955, pp. 326–33.

9. Magnetverstärker-Regelung in Anlagen der Starkstromtechnik. Der Magnetverstärker als Regler, **M. Tschermak, W. Kafka.** *Siemens-Zeitschrift*, Erlangen, Germany, 1955.

10. Anwendungen der Magnetverstärker-Regelung in Anlagen der Starkstromtechnik, **M. Tschermak, W. Kafka.** *Ibid.*

11. Magnetic Amplifiers with Inductive D-C Load, **H. F. Storm.** *AIEE Transactions*, vol. 73, pt. I, Nov. 1954, pp. 466–75.

12. Magnetic Amplifier Circuits, **W. A. Geyger.** McGraw-Hill Book Company, Inc., New York, N. Y., 1954, p. 207.

13. Die Selbsttätige Regelung, **A. Leonhard.** Julius Springer, Berlin, Germany, 1949, p. 188.

Simultaneous Pulled Oscillations in a Triode Oscillator Incorporating Two Oscillatory Circuits

ABD EL-SAMIE MOSTAFA
NONMEMBER AIEE

Synopsis: This paper describes theoretical and experimental investigations of simultaneous pulled oscillations generated in a triode oscillator incorporating two oscillatory circuits. It also includes a physical explanation of pulled oscillations generated in nonlinear oscillatory systems.

A triode oscillator incorporating two oscillatory circuits is capable, under certain conditions, of generating two frequencies simultaneously. The ratio of the two frequencies may be rational or irrational. The oscillations in the latter case are difficult to obtain practically, since they are not self-starting. The important case, therefore, occurs when the two frequencies are related to each other by a ratio q_1/q_2, where q_1 and q_2 are simple integers; this case represents simultaneous locked oscillations, about which several papers have been written[1-3]. In many cases, before locking, the oscillations are pulled to each other, the resulting wave form, representing simultaneous oscillations modulated in both amplitude and frequency, are defined as simultaneous pulled oscillations. To the author's knowledge, no previous investigator analyzed these oscillations. However, pulled oscillations in a single-tuned-circuit oscillator excited by an external signal were previously treated.[4-6]

Whenever there are two or more tuned circuits in a generating system, pulled oscillations may exist. The trouble in operation of an unstrapped multicavity magnetron is ascribed to these oscillations. The hunting phenomena observed when connecting two synchronous machines in parallel is also a form of pulled oscillations. An interesting and important case involving two resonant circuits and exhibiting the present phenomena arises when the load of an oscillator is supplied by a tuned circuit coupled to its tank circuit. Thus a need is felt for a simple analysis of the oscillations.

In a single-tuned-circuit oscillator, the external pulling signal is of constant frequency, i.e., of an ideal stability. However, in the present case, there are two oscillations, any one of which could be considered as the pulling source and the other as the pulled oscillation; the pulling source, i.e., the equivalent signal, is therefore stable neither in frequency nor in amplitude.

Physical Explanation of Pulled Oscillations

BEFORE the theoretical analysis of the present problem, it is interesting to give some physical explanation of pulled oscillations generated in nonlinear oscillatory systems.

When an oscillatory system containing a nonlinear element is acted upon by an external signal, the free oscillation may, under certain conditions, follow the signal exactly, i.e., the frequency of the oscillatory system falls in step with the signal frequency or with a frequency related to it by an integral ratio, keeping the phase angle between the oscillation and the signal constant; these are termed synchronized oscillations. In this locked state, the signal completely controls the oscillation.

In many cases, before locking, the phase angle and the amplitude of oscillation are no longer constant as in the locked state, but slowly varying with time. In this case, the control of the signal decreases but is not lost. The oscillations are pulled to the signal; these are pulled oscillations. They may occur in electrical, mechanical, or electromechanical systems. Examples of the first are electron-tube oscillators, including klystrons, when disturbed by external signals. The oscillations in superregenerative detectors under signal conditions are, in general, pulled oscillations. An example of pulled oscillations in electromechanical systems is the hunting phenomena observed when a synchronous machine is connected to an infinite bus.

In a generating system containing two tuned circuits, there may exist two simultaneous oscillations, any one of which could be considered as an external signal to the other. The present problem is an example of this case. Another example is the hunting phenomena observed when connecting two synchronous machines in parallel.

In all of these cases, the phenomena of pulling can be explained as follows: Consider an electron-tube oscillator acted upon by an external signal of a frequency nearly equal to the free frequency. When the detuning $\Delta\omega/2\pi$ between the free oscillation frequency and the signal frequency is very large, the signal may lose its control and the system operates freely (free oscillation), but there is still an output, although small, of the signal frequency (forced oscillation). The two oscillations, free and forced, will heterodyne resulting in a beat component of a frequency equal to $\Delta\omega/2\pi$. During the interval at which the free and forced oscillations are in phase, the resultant amplitude is maximum. However, at the interval where both oscillations are out of phase, the amplitude of oscillation is minimum. The result is that the amplitude of oscillation will fluctuate between maximum and minimum at a rate equal to the detuning. If $\Delta\omega$ is reduced, but still remains outside the locking range, then during the in-phase part of the beat cycle the signal assists the oscillation, and its control is strong; the signal pulls the oscillation and tries to lock it. At the other part of the beat cycle, the forced oscillation opposes the free one and the control of the signal reduces; the oscillation in this part of the cycle tries to get rid of the signal and to continue freely. The rate at which energy is supplied from the external signal to the oscillatory circuit must accordingly be variable. This variation in instantaneous rate of energy transfer results in disturbing the balance of the phase angle between the oscillation and the applied signal. The variation of the phase angle with time must be such that the instantaneous frequency of oscillation is nearest the signal frequency at the in-phase part of the beat cycle, where the pulling is maximum. Because of pulling, the average frequency of oscillation must be shifted toward the signal frequency; the periodicity of the phase variation, termed the disturbed beat frequency $p/2\pi$, must also be less than the undisturbed beat frequency $\Delta\omega/2\pi$; see equation 11(A). It is evident that $p/2\pi$ must depend on $\Delta\omega$, the amplitude of the signal, and the quality of the oscillatory circuit.

Pulled oscillations represent, therefore, frequency- and amplitude-modulated waves. However, because of the nonlinearity of the tube characteristic, the application of the external signal cannot produce any appreciable increase in the amplitude of oscillation, and thus the degree of amplitude modulation is relatively small.

Although the foregoing explanation is

Paper 57-77, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted January 24, 1956; made available for printing December 20, 1956.

Abd El-Samie Mostafa is with Alexandria University, Alexandria, Egypt.

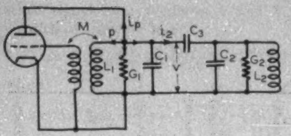


Fig. 1. Tuned-plate oscillator incorporating two oscillatory circuits, capacitively coupled

made for a signal frequency near the free frequency, it can be applied when the signal frequency is near an integral ratio of the free frequency. The explanation is valid whether the applied signal is obtained from an external source or generated in the same oscillatory system; an example of the latter is the simultaneous pulled oscillations of this paper.

It should be noted here that the variations of the phase angle between the applied signal and the superregenerative oscillations, observed in Fig. 6, 7, and 9 of reference 7, are caused by the pulling phenomena. The rate of phase variation is equal to the disturbed beat frequency and must be less than the amount of detuning; the difference reduces rapidly as the detuning increases; see equation 11(A). This is the reason why the rate of phase variation along one quench cycle is very nearly equal to the detuning in quench cycles. It can be stated here, for the first time, that the poor selectivity inherent in superregenerative detectors is caused by the pulling phenomena. It is the pulling that makes the signal able to control the superregenerative oscillations along the whole quench cycle.

Theoretical Analysis

Fig. 1 represents a schematic diagram of a tuned-plate triode oscillator incorporating two oscillatory circuits $L_1-G_1-C_1$ and $L_2-G_2-C_2$ capacitively coupled through C_3. The losses are represented by conductances G_1 and G_2 in order to simplify the analysis, and the results obtained can apply satisfactorily as well to the more usual case where the losses are associated with the inductances. Applying Kirchhoff's first law for the currents meeting at the point P gives the equation

$$C_1 dv/dt + G_1 v + i_2 + i_p + (1/L_1)\int v\,dt = 0 \qquad (1)$$

where i_p is the a-c component of the plate current and v is the a-c component of the voltage across the plate. Other parameters are indicated in this figure.

If the losses are low and the coupling is sufficient, so that the system possesses two resonant peaks, the impedance of the circuit at and near the resonant frequencies is very high compared with its value at other harmonic frequencies. The voltage v can therefore be assumed to be of the form

$$v = V_1 \sin \int(\omega_1 + \phi_{1t}')dt + V_2 \sin \int(\omega_2 + \phi_{2t}')dt = v_1 + v_2 \qquad (2)$$

where

$$v_1 = V_1 \sin \int(\omega_1 + \phi_{1t}')dt$$

and

$$v_2 = V_2 \sin \int(\omega_2 + \phi_{2t}')dt$$

The prime indicates time rate of change. The frequencies $\omega_1/2\pi$ and $\omega_2/2\pi$ are near the resonant frequencies of the oscillatory system and are related so that $\omega_1/\omega_2 = q_1/q_2$, where q_1 and q_2 are simple integers. V_1, V_2, ϕ_{1t}, ϕ_{2t}, ϕ_{1t}', and ϕ_{2t}' are slowly varying functions of time compared with the frequencies of oscillation. When the system is locked, the phase angles ϕ_{1t} and ϕ_{2t} and the amplitudes V_1 and V_2 are constant. Besides, in the locked state, the oscillator is assumed to develop its amplitudes V_1 and V_2 up to the saturation values determined by its nonlinear characteristic. Under a pulling state, there is practically no further increase in amplitude. V_1 and V_2 could therefore be considered constant.

In the following analysis, assume that the variation in each of V_1 and V_2 is very small and neglected. This assumption is experimentally satisfied in almost all cases. Using the analysis given in reference 8, and aided by Fig. 1, it is clear that the current i_2 is given by the following equation

$$\begin{aligned} i_2 = v_1\{[Yj(\omega_1+\phi_{1t}') - (j/2)(d\phi_{1t}'/dt)\times \\ d^2 Yj(\omega_1+\phi_{1t}')/d(\omega_1+\phi_{1t}')^2 + \ldots]\} + \\ v_2\{[Yj(\omega_2+\phi_{2t}') - (j/2)(d\phi_{2t}'/dt)\times \\ d^2 Yj(\omega_2+\phi_{2t}')/d(\omega_2+\phi_{2t}')^2 + \ldots]\} \end{aligned}$$

where $Yj(\omega+\phi_t')$ is the admittance of the circuit at an angular frequency $(\omega+\phi_t')$ and is given by

$$\begin{aligned} Yj(\omega+\phi_t') = \{[j(\omega+j\phi_t')C_2 + \\ 1/j(\omega+\phi_t')L_2]/[1+C_2/C_3 - \\ 1/(\omega+\phi_t')^2 L_2 C_3]\} + G_2/[1+C_2/C_3 - \\ 1/(\omega+\phi_t')^2 L_2 C_3]^2 \end{aligned}$$

However, ϕ_t is slowly varying with time and thus the amplitude of ϕ_t' is very small compared with ω. The quasi-steady-state method of analysis can therefore be used, giving

$$i_2 \approx v_1[Yj(\omega_1+\phi_{1t}')] + v_2[Yj(\omega_2+\phi_{2t}')] \qquad (3)$$

Substitution in equation 1 gives

$$C_{eq}(\omega)dv/dt + G_{eq}(\omega)v + i_p + (1/L_1)\int v\,dt = 0 \qquad (4)$$

where $C_{eq}(\omega)$ and $G_{eq}(\omega)$ are equivalent capacitance and conductance respectively and are functions of frequency. At a frequency $(\omega_1+\phi_{1t}')/2\pi$, and neglecting terms containing G_2^2, they are given by

$$C_{eq} = C_1 + \{[C_2 - 1/(\omega_1+\phi_{1t}')^2 L_2]/[1+C_2/C_3 - 1/(\omega_1+\phi_{1t}')^2 L_2 C_3]\}$$

$$G_{eq} = G_1 + \{G_2/[1+C_2/C_3 - 1/(\omega_1+\phi_{1t}')^2 L_2 C_3^2]\}$$

At a frequency $(\omega_2+\phi_{2t}')/2\pi$, the same expressions are used, but $(\omega_1+\phi_{1t}')$ is replaced by $(\omega_2+\phi_{2t}')$. The a-c component of the plate current can be represented as a power series of v, thus

$$i_p = \sum_{n=1} a_n v^n$$

The upper limit of the summation depends on the number of terms required

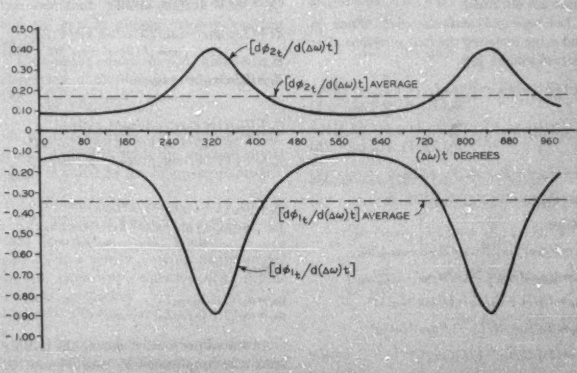


Fig. 2. Rate of phase variations $d\phi_{1t}/d(\Delta\omega)t$ and $d\phi_{2t}/d(\Delta\omega)t$ of pulled oscillations

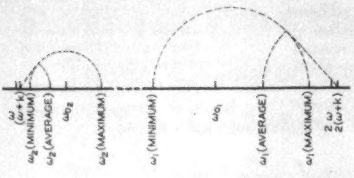


Fig. 3. Frequency relations of pulled oscillations, Z_{1l} is positive

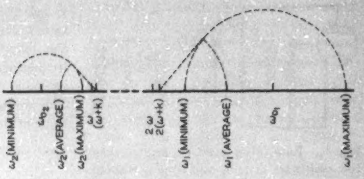


Fig. 4 Frequency relations of pulled oscillations, Z_{1l} is negative

in the expansion for satisfactory representation, but generally it is infinity, and v is given by equation 2. Differentiating v once and twice and substituting in equation 4 gives an equation containing the following variables; ϕ_{1t}', ϕ_{1t}'', $\phi_{1t}'^2$, ϕ_{2t}', ϕ_{2t}'', $\phi_{2t}'^2$, and

$$\begin{matrix}\sin\\ \cos\end{matrix}\int(\omega_1+\phi_{1t}')dt,\quad \begin{matrix}\sin\\ \cos\end{matrix}\int(\omega_2+\phi_{2t}')dt,$$

$$\begin{matrix}\sin\\ \cos\end{matrix}\int n(\omega_1+\phi_{1t}')dt,\quad \begin{matrix}\sin\\ \cos\end{matrix}\int m(\omega_2+$$

$$\phi_{2t}')dt,\quad \text{and}\ \begin{matrix}\sin\\ \cos\end{matrix}\int[n(\omega_1+\phi_{1t}')\pm$$

$$m(\omega_2+\phi_{2t}')]d$$

where n and m are integers.

Usually ϕ'_{1t} and ϕ'_{2t} are very slowly varying compared with ω_1 and ω_2 so that ϕ_{1t}'', ϕ_{2t}'', $\phi_{1t}'^2$, and $\phi_{2t}'^2$ can be neglected; this means that the quasi-steady-state method can be used. At the same time, ϕ_{1t}' and ϕ_{2t}' could be satisfactorily considered constant for an interval containing several cycles of oscillation. For this small time interval, coefficients of

$$\cos\int(\omega_1+\phi_{1t}')dt,\ \cos\int(\omega_2+\phi_{2t}')dt,$$
$$\sin\int(\omega_1+\phi_{1t}')dt,\ \text{and}\ \sin\int(\omega_2+\phi_{2t}')dt$$

could be separately equated with zero. By carrying out that process, four relations are obtained.

Let $\omega_1=q_1\omega$ and $\omega_2=q_2\omega$; since q_1 and q_2 are integers, the four relations after rearrangement are

$$Z_1+(2/q_1\omega)\phi_{1t}'=[-q_1\omega L_1/(1+L_1/\alpha_1L_2)]\times F_{p_1}(V_1,\ V_2,\ \phi_{1t},\ \phi_{2t}) \quad (5)$$

$$Z_2+(2/q_2\omega)\phi_{2t}'=[-q_2\omega L_1/(1+L_1/\alpha_2L_2)]\times F_{q_1}(V_1,\ V_2,\ \phi_{1t},\ \phi_{2t}) \quad (6)$$

$$G_1+G_{2p}=-F_{p_2}(V_1,\ V_2,\ \phi_{1t},\ \phi_{2t}) \quad (7)$$

$$G_1+G_{2q}=-F_{q_2}(V_1,\ V_2,\ \phi_{1t},\ \phi_{2t}) \quad (8)$$

where

$$Z_1=[(q_1\omega)^2/\omega_{01}^2]-1\approx 2(q_1\omega-\omega_{01})/\omega_{01}$$

$$Z_2=[(q_2\omega)^2/\omega_{02}^2]-1\approx 2(q_2\omega-\omega_{02})/\omega_{02}$$

$$G_{2p}\approx G_2/[1+C_2/C_3-1/(q_1\omega)^2L_2C_3]^2$$

$$G_{2q}\approx G_2/[1+C_2/C_3-1/(q_2\omega)^2L_2C_3]^2$$

$$\alpha_1=(1+C_2/C_3-1/L_2C_3\omega_{01}^2)^2$$

$$\alpha_2=(1+C_2/C_3-1/L_2C_3\omega_{02}^2)^2$$

F_{p_1}, F_{q_1}, F_{p_2}, and F_{q_2} are functions of V_1, V_2, ϕ_{1t}, and ϕ_{2t}; they are given in Appendix I. ω_{01} and ω_{02} are the resonant angular frequencies of the system, in the absence of the electron tube. Neglecting losses, they are given by

$$\omega_{01}^2=H+\sqrt{H^2-\frac{(1/L_1C_1)(1/L_2C_2)}{1+C_3/C_1+C_3/C_2}}$$

where

$$H=\frac{(1/L_1C_1)(1+C_3/C_2)+(1/L_2C_2)(1+C_3/C_1)}{2(1+C_3/C_1+C_3/C_2)}$$

The same expression is used for ω_{02}^2, but the sign before the radical is negative. For many practical cases, equations 5 through 8 satisfactorily give the general amplitude-phase-frequency relations, at any time t, of a triode oscillator incorporating two oscillatory circuits.

To simplify the analysis without losing generality, only the first three terms of the plate-current series are considered. Again, the case where $q_1:q_2=2:1$ is carried out. In this way, a satisfactory demonstration for the general behavior of the circuit can be obtained simply. Substituting $q_1=2$ and $q_2=1$ in equations 5 through 8 and using parameters up to a_3 only, gives

$$Z_1+(1/\omega)\phi_{1t}'=(a_{2_1}/2)(2\omega L_1)(V_2^2/V_1)\times \cos\int(2\phi_{2t}'-\phi_{1t}')dt \quad (5A)$$

$$Z_2+(2/\omega)\phi_{2t}'=-a_{2_2}\omega L_1V_1\times \cos\int(2\phi_{2t}'-\phi_{1t}')dt \quad (6A)$$

$$G_1+[G_2/(1+C_2/C_3-1/4\omega^2L_2C_3)^2]+[a_1+(3/4)a_3(V_1^2+2V_2^2)+(a_2/2)\times(V_2^2/V_1)\sin\int(2\phi_{2t}'-\phi_{1t}')dt]=0 \quad (7A)$$

$$G_1+[G_2/(1+C_2/C_3-1/\omega^2L_2C_3)^2]+[a_1+(3/4)a_3(V_2^2+2V_1^2)+a_2V_1\times \sin\int(2\phi_{2t}'-\phi_{1t}')dt]=0 \quad (8A)$$

where

$$a_{2_1}=a_2/(1+L_1/\alpha_1L_2)$$
$$a_{2_2}=a_2/(1+L_1/\alpha_2L_2)$$

As stated previously, during the pulling state the amplitudes V_1 and V_2 can be considered constant; in this way, equations 5(A) and 6(A) give satisfactory information about pulled oscillations. At the locking state $\phi_{1t}'=0=\phi_{2t}'$, and the phase angles ϕ_{1t} and ϕ_{2t} are no longer variables but are constants, and are given by ϕ_1 and ϕ_2 respectively. Substitution in equations 5(A) and 6(A) gives

$$Z_1=(a_{2_1}/2)(2\omega L_1)(V_2^2/V_1)\cos(2\phi_2-\phi_1)$$

$$Z_2=-a_{2_2}\omega L_1V_1\cos(2\phi_2-\phi_1)$$

The locking range is obtained when $\cos(2\phi_2-\phi_1)=\pm 1$, giving

$$(a_{2_1}/2)2\omega L_1V_2^2/V_1=Z_{1l} \quad (9A)$$

$$-a_{2_2}\omega L_1V_1=Z_{2l} \quad (9B)$$

where Z_{1l} and Z_{2l} are the values of Z_1 and Z_2 at the limit of the locking range. Substitution in equations 5(A), 6(A) and 9 and rearrangement gives

$$d\theta_t/dt=\Delta\omega-(\Delta\omega)_l\cos\theta_t \quad (10)$$

where

$$'=2\phi_{2t}-\phi_{1t}$$

$$\Delta\omega=\omega(Z_1-Z_2)=(2\omega^2/\omega_{01}\omega_{02})(2\omega_{02}-\omega_{01})\approx 2\omega_{02}-\omega_{01}$$

and

$$(\Delta\omega)_l=\text{locking range}=\omega(Z_{1l}-Z_{2l})=a_{2_2}\omega^2L_1V_1[1+(a_{2_1}/a_{2_2})(V_2^2/V_1^2)]=2\omega_{02l}-\omega_{01l}$$

The index l indicates that the values are at the limit of the locking range. Equation 10 is similar to the phase-frequency relation of a single-tuned-circuit oscillator disturbed by an ex-

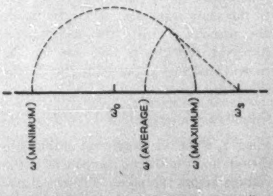


Fig. 5. Frequency relations of pulled oscillations in a single-tuned-circuit oscillator

ternal signal; see references 4 and 6. Solution of equation 10 is

$$\theta_t = 2\tan^{-1}[\tan(\beta/2)\tan((pt-\beta)/2)] \quad (11)$$

where

$$\cos\beta = (\Delta\omega)_l/\Delta\omega \text{ and } p = \sqrt{(\Delta\omega)^2-(\Delta\omega)_l^2} \quad (11A)$$

It is clear that the angle β varies from zero (at the limit of the locking range) to $\pi/2$ at a detuning very far from locking. The frequency $p/2\pi$ may be defined as the modulation frequency; it is the disturbed beat frequency and is less than the undisturbed beat frequency $\Delta\omega/2\pi$. This reduction, caused by pulling, reduces as the detuning $\Delta\omega/2\pi$ increases.

Substituting equation 11 in equation 10 and rearrangement gives

$$d\theta_t/dt = 2\phi_{2t}'-\phi_{1t}' = \Delta\omega\sin^2\beta/[1+\cos\beta\cos(pt-\beta)] = p+p\sin\beta\times \sum_{m=1}^{\infty}(-1)^m\left[\sum_{\substack{n=m,\ m+2,\\ m+4,\dots}}(\cos^n\beta/2^{n-1})^m C_{(n-m)/2}\right]\cos m(pt-\beta) \quad (12)$$

where

$$^mC_{(n-m)/2} = \{m(m-1)(m-2)(m-3)\dots[m-(n-m)/2+1]\}/[(n-m)/2]!$$

Equation 12 is valid, if $\cos\beta$ is less than and not equal to unity, i.e., if the system is in the pulled state.

Inside the locking range, $\Delta\omega<(\Delta\omega)_l$ and both β and p are imaginary. Let them be given by $j\beta_1$ and jp_1; therefore,

$$\theta_t = 2\tan^{-1}[-\tanh(\beta_1/2)\times\tanh((p_1t-\beta_1)/2)]$$

With the passage of time, $\tanh((p_1t-\beta_1)/2)$ goes asymptotically toward unity, and θ_t goes toward a constant value θ, given by $\cos\theta=\Delta\omega/(\Delta\omega)_l$; this result may be obtained directly from equation 10 by letting $d\theta_t/dt=0$ and $\theta_t=\theta$.

It is interesting to compare the phase-frequency relations given by equations 5(A) and 6(A) with the corresponding relations of the second-harmonic and second-subharmonic oscillations, respectively, of a single-tuned-circuit oscillator with an applied external signal in its grid circuit. If the applied signal in a single-tuned-circuit triode oscillator is given by $e_s=E\sin(\omega_st+\phi_2)$, and the plate circuit tuned to a frequency $\omega_{0_a}/2\pi$ near $2\omega_s/2\pi$, an output second-harmonic voltage

$$v = V_1\sin\int(2\omega_s+\phi_{1t}')dt$$

may be generated and the phase-frequency relation is given in Appendix II as the equation which follows

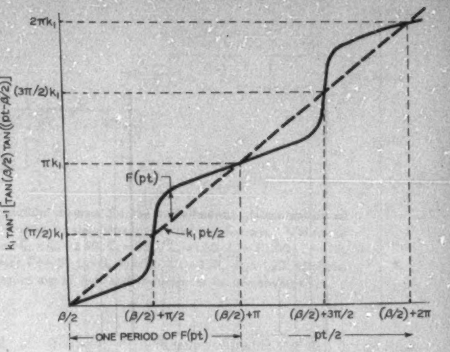


Fig. 6. Representation of $k_1\tan^{-1}\{\tan(\beta/2)\tan[(pt-\beta)/2]\}$

$$Z_1+(1/\omega_{0_a})d\phi_{1t}/dt = (1/4\omega_{0_a}C)[\mu^2E^2/V_1(1-\mu M/L)^2]b_2(1-\mu M/L)^2\times\cos(2\phi_2-\phi_{1t}) \quad (13)$$

where

$$Z_1 = 2(2\omega_s-\omega_{0_a})/\omega_{0_a}$$

However, if $e_s=E\sin(\omega_st+\phi_1)$, and the plate circuit is tuned to a frequency $\omega_{0_b}/2\pi$ near $\omega_s/4\pi$, an output second-subharmonic voltage

$$v = V_2\sin\int[(\omega_s/2)+\phi_{2t}']dt$$

may be generated and the phase-frequency relation is also given in Appendix II as

$$Z_2+(2/\omega_{0b})d\phi_{2t}/dt = -(2/\omega_{0b}C)\times[\mu E/(\mu M/L-1)]b_2(1-\mu M/L)^2\times\cos(2\phi_{2t}-\phi_1) \quad (14)$$

where

$$Z_2 = 2[(\omega_s/2)-\omega_{0b}]/\omega_{0b}$$

In both equations 13 and 14, ϕ_1 and ϕ_2 are constant phase angles while ϕ_{1t} and ϕ_{2t} are, in general, slowly varying angles.

Comparing equations 5(A) and 13, it is clear that 5(A) represents an oscillation

$$v = V_1\sin\int(2\omega+\phi_{1t}')dt$$

of a hypothetical single-tuned-circuit triode oscillator with an applied signal which is equal to the following relation

$$V_2[(\mu M/L-1)/\mu]\times\sqrt{4\omega^2L_1C/(1+L_1/\alpha_1L_2)}\sin\int(\omega+\phi_{2t}')dt$$

and ϕ_{2t} is a slowly varying phase angle. In the same way, equation 6(A) represents an oscillation

$$v = V_2\sin\int(\omega+\phi_{2t}')dt$$

of a hypothetical single-tuned-circuit triode oscillator with an applied signal which is equal to the following relation

$$V_1[(\mu M/L-1)/2\mu][\omega^2L_1C/(1+L_1/\alpha_2L_2)]\times\sin\int(2\omega+\phi_{1t}')dt$$

and ϕ_{1t} is also a slowly varying phase angle.

It is concluded therefore that the two pulled oscillations, generated in a triode oscillator incorporating two oscillatory circuits, could be considered as being generated in two separate single-tuned-circuit oscillators, the output voltage of either oscillator is applied as an external signal to the other. In the locked state, where the amplitudes and phase angles are constant, the system could be considered as two simple synchronized oscillators in the subharmonic or harmonic locked state.

DETERMINATION OF FREQUENCY SPECTRA OF PULLED OSCILLATIONS

The use of equations 5(A), 6(A), 9(A), 9(B), and 12 gives

$$d\phi_{1t}/dt = 2k-\{\sin\beta\tan\beta/[1+\cos\beta\times\cos(pt-\beta)]\}\omega Z_{1l} = 2k-\omega Z_{1l}[p/(\Delta\omega)_l]\left\{1+\sin\beta\sum_{m=1}(-1)^m\times\left[\sum_{\substack{n=m,\ m+2,\\ m+4,\dots}}(\cos^n\beta/2^{n-1})^m C_{(n-m)/2}\right]\times\cos m(pt-\beta)\right\} \quad (15)$$

$$d\phi_{2t}/dt = k-\{\sin\beta\tan\beta/[1+\cos\beta\times\cos(pt-\beta)]\}\omega Z_{2l}/2 = k-(\omega Z_{2l}/2)\times[p/(\Delta\omega)_l]\left\{1+\sin\beta\sum_{m=1}(-1)^m\times\left[\sum_{\substack{n=m,\ m+2,\\ m+4,\dots}}(\cos^n\beta/2^{n-1})^m C_{(n-m)/2}\right]\times\cos m(pt-\beta)\right\} \quad (16)$$

where

$$k = (\omega^2/2)(Z_1Z_{2l}-Z_2Z_{1l})/(\Delta\omega)_l$$

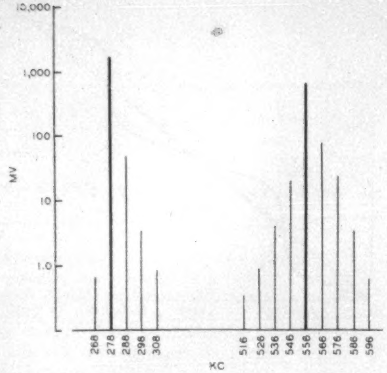


Fig. 7. Frequency spectrum of pulled oscillations; ordinate represents values proportional to the amplitudes of the different frequency components; $q_1:q_2=2:1=556$ kc:278 kc

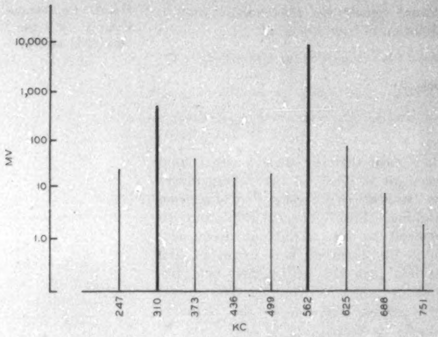


Fig. 9. Same as Fig. 6, but $q_1:q_2=9:5\approx562$ kc: 310 kc

It is clear from equations 15 and 16 that the instantaneous angular frequencies $(2\omega+d\phi_{1l}/dt)$ and $(\omega+d\phi_{2l}/dt)$ vary periodically with the same fundamental frequency $p/2\pi$ which reduces as the amount of detuning approaches the locking value. The wave forms of the instantaneous frequencies are of the same shape, assuring that the pulling phenomena is caused by the continual disturbance of the instantaneous rate of energy transfer between the two individual locked oscillations; see Fig. 2.

Taking $\Delta\omega/(\Delta\omega)_l=1.385$, $\omega Z_{1l}/\Delta\omega=0.396$, $Z_{1l}/Z_{2l}=-1.21$, and $k/\Delta\omega=0.02$, both $d\phi_{1l}/d(\Delta\omega)t$ and $d\phi_{2l}/d(\Delta\omega)t$ are calculated from equations 15 and 16 and drawn in Fig. 2. It is also clear from equations 15 and 16 that the average angular frequencies of oscillation are given by the equations which follow

$$\omega_1\,(\text{average})=2\omega+(d\phi_{1l}/dt)_{\text{av}}=2\{\omega+k-[p/2(\Delta\omega)_l]\omega Z_{1l}\}$$

$$\omega_2\,(\text{average})=\omega+(d\phi_{2l}/dt)_{\text{av}}=\{\omega+k-[p/2(\Delta\omega)_l]\omega Z_{2l}\}$$

The average frequencies of oscillation are therefore not in the ratio of 2 to 1. However, just outside the locking range, p is very small and they are approximately in the ratio of 2 to 1. Maximum and minimum values of the instantaneous angular frequencies are,

$$\omega_1\,(\text{maximum})=2\left[(\omega+k)-\frac{\Delta\omega-(\Delta\omega)_l}{(\Delta\omega)_l}(\omega Z_{1l}/2)\right],\ \omega_1\,(\text{minimum})=2\left[(\omega+k)-\frac{\Delta\omega+(\Delta\omega)_l}{(\Delta\omega)_l}(\omega Z_{1l}/2)\right]$$

$$\omega_2\,(\text{maximum})=\left[(\omega+k)-\frac{\Delta\omega+(\Delta\omega)_l}{(\Delta\omega)_l}(\omega Z_{2l}/2)\right],\ \omega_2\,(\text{minimum})=\left[(\omega+k)-\frac{\Delta\omega-(\Delta\omega)_l}{(\Delta\omega)_l}(\omega Z_{2l}/2)\right]$$

It should be noted that Z_{1l} and Z_{2l} are of opposite signs, see equations 9(A) and 9(B), and thus Z_1 and Z_2 are of opposite signs; if Z_{1l} is positive, Z_{2l} is negative and vice versa. The given expressions of maximum and minimum angular frequencies are for positive values of Z_{1l}; for negative values, the maximum and minimum are to be interchanged.

It is evident from the given expressions of the average, maximum, and minimum angular frequencies that as the system goes toward locking (i.e., as $\Delta\omega$ approaches $(\Delta\omega)_l$), ω_1 (maximum), ω_1 (average), ω_2 (minimum), and ω_2 (average) approach the corresponding locked angular frequencies 2ω and ω respectively, while the deviation of ω_1 (minimum) and ω_2 (maximum) from ω_1 (average) and ω_2 (average), and thus from 2ω and ω respectively, increases; see Fig. 3 which is drawn for a detuning near the locked state and for Z_{1l} positive. It should be noted here that the locked frequency could be considered as the frequency of the equivalent signal. At the limit of the locking range $d\phi_{1l}/dt=0=d\phi_{2l}/dt$, and the maximum and minimum angular frequencies must coincide with the corresponding locked frequencies. The result is that, as the system goes toward locking, ω_1 (maximum) and ω_2 (minimum) gradually approach 2ω and ω respectively, while the deviation of ω_1 (minimum) and ω_2 (maximum) from 2ω and ω increases;

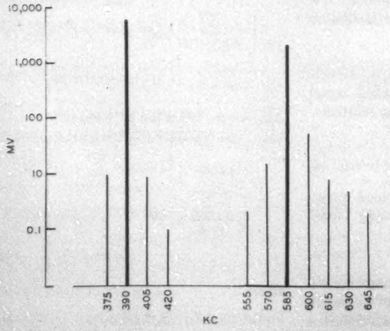


Fig. 8. Same as Fig. 6, but $q_1:q_2=3:2=585$ kc: 390 kc

but at the instant of locking both ω_1 (minimum) and ω_2 (maximum) jump suddenly to 2ω and ω respectively. This conclusion is valid for all types of pulled oscillations. If Z_{11} is negative, the same phenomena occurs but ω_1 (minimum) and ω_1 (maximum) are interchangeable and thus ω_2 (maximum) and ω_2 (minimum) are also interchangeable; see Fig. 4.

Again, aided with the given expressions of the maximum, minimum, and average angular frequencies, the following conclusions are obtained for Z_{11} positive. Both ω_1 (maximum) and ω_1 (minimum) are lower than the corresponding angular frequency 2ω at the locked state, while ω_2 (maximum) and ω_2 (minimum) are higher than ω. The average angular frequency ω_1 (average) is equal to the geometrical mean of $[2(\omega+k)-\omega_1$ (maximum)] and $[2(\omega+k)-\omega_1(\text{minimum})]$, while ω_2 (average) is equal to the geometrical mean of $[\omega_2$ (maximum) $-(\omega+k)]$ and $[\omega_2$ (minimum) $-(\omega+k)]$; this characteristic is used in the graphical construction of Fig. 3. The free angular frequency ω_{01} is lower than the corresponding average frequency ω_1 (average), while ω_{02} is higher than ω_2 (average). The average frequency is not equal to the corresponding free frequency, but it is shifted (pulled) away from the free frequency toward the locked frequency, i.e., toward the equivalent signal frequency. The allocations of the different angular frequencies of the system are shown in Fig. 3 for Z_{11} positive, and in Fig. 4 for Z_{11} negative. It should be noted here that Fig. 3 corresponds to Fig. 2. Figs. 3 and 4 can be used to clarify the mechanism of pulling. It is clear from these figures that the instantaneous angular frequency, being variable between ω (maximum) and ω (minimum), is always lower or higher than the corresponding locked frequency, depending on whether the free frequency is lower or higher than the locked frequency. This means that during the pulled state the instantaneous frequency of oscillation is never equal to the locked frequency, and thus the instantaneous rate of energy transfer is continually disturbed. The contribution of the equivalent signal (of a frequency equal to the locked frequency) to the rate of energy transfer is expected to favor any oscillation of a frequency nearer to its own and thence the pulling of the average frequency away from the free frequency toward the locked frequency.

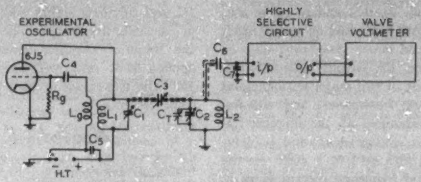


Fig. 10. Connection diagram for the experimental determination of the frequency spectra of simultaneous pulled oscillations. Values of components are: $C_1=C_2=150$, $C_3=350$, $C_4=160$, $C_5=1{,}000$, $C_6=10$, $C_7=100$, trimmer $C_T=5$, $L_1=L_2=460$, $L_g=350$, $R_g=120$ kilohms; capacitances are in $\mu\mu f$, and inductances in microhenrys

The frequency allocations, corresponding to Fig. 3, of pulled oscillations in a single-tuned-circuit oscillator disturbed by an external signal of frequency $\omega_s/2\pi$ near its free frequency $\omega_0/2\pi$, is shown in Fig. 5; here again the average angular frequency of oscillation ω (average) is pulled away from ω_0 toward ω_s.

The voltage appearing across the network for the pulled oscillations is given by

$$\begin{aligned} v = {} & V_1 \sin \int 2\{(\omega+k)-(\omega Z_{11}/2)[\sin\beta\times \\ & \tan\beta/(1+\cos\beta\cos(pt-\beta)]\}\,dt+ \\ & V_2 \sin \int \{(\omega+k)-(\omega Z_{21}/2)[\sin\beta\times \\ & \tan\beta/(1+\cos\beta\cos(pt-\beta))]\}\,dt \\ & = V_1 \sin 2\,[(\omega+k)t-\phi_{1t_0}]+ \\ & \qquad V_2 \sin\,[(\omega+k)t-\phi_{2t_0}] \qquad (17) \end{aligned}$$

where

$$\phi_{1t_0}=[\omega Z_{11}/(\Delta\omega)_l]\tan^{-1}[(\tan(\beta/2))\times \tan((pt-\beta)/2)]$$

$$\phi_{2t_0}=[\omega Z_{21}/(\Delta\omega)_l]\tan^{-1}[(\tan(\beta/2))\times \tan((pt-\beta)/2)]$$

Either ϕ_{1t_0} or ϕ_{2t_0} is of the form

$$k_1\tan^{-1}[(\tan(\beta/2))\tan((pt-\beta)/2)]$$

where

$$k_1=\omega Z_{11}/(\Delta\omega)_l \text{ for } \phi_{1t_0}$$

and

$$k_1=\omega Z_{21}/(\Delta\omega)_l \text{ for } \phi_{2t_0}$$

It is interesting to discuss the behavior of the function

$$k_1\tan^{-1}[(\tan(\beta/2))\tan((pt-\beta)/2)]$$

As the time t continues, $pt/2$ will pass through $\beta/2$, $[(\beta/2)+\pi/2]$, $[(\beta/2)+\pi]$, $[(\beta/2)+(3\pi/2)]$, $[(\beta/2)+2\pi]$, etc., and the tangent will become 0, $+\infty$, -0, $-\infty$, 0, etc., in succession; at these instants the function will be equal to 0, $(\pi/2)k_1$, πk_1, $(3\pi/2)k_1$, $2\pi k_1$, etc. The curve passing through these points is the dashed straight line shown in Fig. 6. At the other instants, the function will fluctuate around this straight line. The fluctuations are repeated with a periodicity equal to $p/2\pi$. The function

$$k_1\tan^{-1}[(\tan(\beta/2))\tan((pt-\beta)/2)]$$

is therefore equal to the sum of two functions. One is linearly related with time and equal to $k_1pt/2$, and the other is a periodic function of fundamental frequency equal to $p/2\pi$ and denoted by $F(pt)$; see Fig. 6. The analytical expression of

$$k_1\tan^{-1}[(\tan(\beta/2))\tan((pt-\beta)/2)]$$

is the integration of equation 12 after being multiplied by $k_1/2$. Thus,

$$\begin{aligned} & k_1\tan^{-1}[(\tan(\beta/2))\tan((pt-\beta)/2)] \\ & \quad =(k_1pt/2)+(k_1/2)\sin\beta\times \\ & \quad \sum_{m=1}^{\infty}(-1)^m(1/m)\Bigg[\sum_{\substack{n=m,\,m+2,\\ m+4,\,\ldots}}\times \\ & \quad (\cos^n\beta/2^{n-1})^m C_{(n-m)/2}\Bigg]\times \\ & \quad \sin m(pt-\beta)=k_1pt/2+F(pt) \qquad (18) \end{aligned}$$

Equation 17 can therefore be written as

$$\begin{aligned} v = {} & V_1\sin\{2[\omega+k-(\omega Z_{11}/(\Delta\omega)_l)(p/2)]t- \\ & F_1(pt)\}+V_2\sin\{[\omega+k- \\ & (\omega Z_{21}/(\Delta\omega)_l)\times(p/2)]t- \\ & (2Z_{21}/Z_{11})F_1(pt)\} \qquad (19) \end{aligned}$$

$$\begin{aligned} v = {} & V_1\{\sin[\omega_1(\text{average})\,t]\cos F_1(pt)- \\ & \cos[\omega_1(\text{average})\,t]\sin F_1(pt)\}+V_2\times \\ & \{\sin[\omega_2(\text{average})\,t]\cos[(2Z_{21}/Z_{11})\times \\ & F_1(pt)]-\cos[\omega_2(\text{average})\,t]\sin \\ & [(2Z_{21}/Z_{11})F_1(pt)]\} \qquad (20) \end{aligned}$$

where $F_1(pt)$ is equal to $F(pt)$, after replacing k_1 by $[\omega Z_{11}/(\Delta\omega)_l]$, it is represented as

$$F_1(pt)=b_1\sin(pt-\beta)+b_2\sin 2(pt-\beta)+ b_3\sin 3(pt-\beta)+\ldots$$

where b_1, b_2, b_3, etc., are constants depending on $[\omega Z_{11}/(\Delta\omega)_l]$ and β; see equation 18. It is therefore clear that both $\cos[F_1(pt)]$ and $\cos[(2Z_{21}/Z_{11})F_1(pt)]$ are periodic functions of the form

$$d_0+d_1\cos(pt-\beta)+d_2\cos 2(pt-\beta)+ d_3\cos 3(pt-\beta)+\ldots$$

Also, both $\sin F_1(pt)$ and $\sin[(2Z_{21}/Z_{11})\times F_1(pt)]$ are periodic functions of the form

$$g_1\sin(pt-\beta)+g_2\sin 2(pt-\beta)+ g_3\sin 3(pt-\beta)+\ldots$$

where d_0, d_1, d_2, etc., and g_1, g_2, g_3, etc., are constants depending on $[\omega Z_{1t}/(\Delta\omega)_t]$, (Z_{2t}/Z_{1t}), and β.

Therefore, the output pulled oscillations consist of two major resonances at the average frequencies, and side-frequency resonances at the angular frequencies $[\omega_1(\text{average})\pm np]$ and $[\omega_2(\text{average})\pm np]$, with $n=1, 2, 3$, etc.; see the experimental frequency spectra given in Figs. 7 through 9. If the system is just outside the locking range, p is very small and the two average frequencies are approximately in the integral ratio as if the pulled oscillations were locked; see Figs. 7 and 8. However, if the detuning is not near the locking range, the two average frequencies differ slightly from that ratio; see Fig. 9.

Experimental Verification

To test the validity of the foregoing theoretical relations regarding the frequency spectra of pulled oscillations in a triode oscillator incorporating two oscillatory circuits, the following experimental work was carried out.

The oscillator used is of the ordinary tuned-plate type, the load of which is supplied by a tuned circuit capacitively coupled to the plate tank circuit; see Fig. 10. The grid and plate coils are well shielded; the coil L_2 is also properly shielded in order to avoid any inductive coupling between L_2 and any of the grid and plate coils. The only coupling between the two tuned circuits is through the calibrated variable air capacitor C_3, as stated before in the theory. Both the rotor and stator of the capacitor C_3 are properly isolated. Capacitor C_3 is also highly shielded, and is connected to both C_1 and C_2 through short and shielded leads in order to reduce any inductance between C_3 and either C_1 or C_2. Both C_1 and C_2 are calibrated variable air capacitors each of 150-$\mu\mu f$ (micromicrofarad) maximum value, while the maximum value of C_3 is 350 $\mu\mu f$.

The experimental values of the quality factor Q, of either of the coils L_1 or L_2, with its associated shield are not low. Q varies from 45 to 30 in most of the operating frequency range, and thus the approximations given in the theoretical analysis are satisfactory, although Q was assumed constant. The values of the coefficient of coupling used in the experiments are sufficient, being variable from 0.45 to 0.6.

The experimental frequency spectra are obtained by applying a part of the output voltage of the experimental oscillator, through a capacitance divider (C_6 and C_7), to a highly selective circuit of variable frequency. The output of the frequency-selective circuit is then applied to a valve voltmeter; see Fig. 10. The experimental curves are taken from reference 2.

Appendix I

$$\begin{aligned}
F_{P_1}=\sum&(a_n/2^{n-1})V_1^{\,n-m-1}V_2^{\,m}\,{}^{n}C_m{}^{n-m}C_{r_1}\\
&{}^{m}C_{r_2}(\{[j^{(-n+2r_1+2r_2)}\cos((m-2r_2)\times\\
&\phi_{2t}+(n-m-2r_1\mp1)\times\\
&\phi_{1t})]_{n-m-2r_1\mp1=(2r_2-m)q_2/q_1}+\\
&[j^{(-n-2m+2r_1-2r_2)}\cos((m-2r_2)\phi_{2t}-\\
&(n-m-2r_1\mp1)\phi_{1t})]_{n-m-2r_1\mp1=-(2r_2-}\\
&{}_{m)q_2/q_1}\}_{n\text{-even}}+\{[j^{1-n+2r_1+2r_2}\times\\
&\sin((m-2r_2)\phi_{2t}+(n-m-2r_1\mp1)\times\\
&\phi_{1t})]_{n-m-2r_1\mp1=(2r_2-m)q_2/q_1}-\\
&[j^{1-n-2m+2r_1-2r_2}\sin((m-2r_2)\phi_{2t}-\\
&(n-m-2r_1\mp1)\times\\
&\qquad\phi_{1t})]_{n-m-2r_1\mp1=-(2r_2-m)q_2/q_1}\}_{n\text{-odd}})
\end{aligned}$$

$$\begin{aligned}
F_{q_1}=\sum&(a_n/2^{n-1})V_1^{\,n-m-1}V_2^{\,m}\,{}^{n}C_m{}^{n-m}C_{r_1}\times\\
&{}^{m}C_{r_2}(\{[j^{-n+2r_1+2r_2}\cos((n-m-\\
&2r_1)\phi_{1t}+(m-2r_2\mp1)\times\\
&\phi_{2t})]_{n-m-2r_1=(2r_2-m\pm1)q_2/q_1}+\\
&[j^{-n-2m+2r_1-2r_2}\cos((n-m-\\
&2r_1)\phi_{1t}+(m-2r_2\mp1)\times\\
&\phi_{2t})]_{n-m-2r_1=(m-2r_2\mp1)q_2/q_1}\}_{n\text{-even}}+\\
&\{[j^{1-n+2r_1+2r_2}\sin((n-m-2r_1)\phi_{1t}+\\
&(m-2r_2\mp1)\phi_{2t})]_{n-m-2r_1=(2r_2-}\\
&{}_{m\pm1)q_2/q_1}-[j^{1-n+2m+2r_1-2r_2}\sin(-(n-\\
&m-2r_1)\phi_{1t}+(m-2r_2\mp1)\times\\
&\qquad\phi_{2t})]_{n-m-2r_1=(m-2r_2\mp1)q_2/q_1}\}_{n\text{-odd}})
\end{aligned}$$

$$\begin{aligned}
F_{p_2}=\sum&(a_n/2^{n-1})V_1^{\,n-m-1}V_2^{\,m}\,{}^{n}C_m{}^{n-m}C_{r_1}\times\\
&{}^{m}C_{r_2}(\{\mp[j^{-n-2r_1+2r_2}\times\\
&\sin((m-2r_2)\phi_{2t}+(n-m-2r_1\mp1)\times\\
&\phi_{1t})]_{n-m-2r_1\mp1=(2r_2-m)q_2/q_1}\pm\\
&[j^{-n-2r_1+2r_2}\sin((m-2r_2)\phi_{2t}+(n-\\
&m-2r_1\mp1)\phi_{1t})]_{n-m-2r_1\mp1=}\\
&{}_{-(2r_2-m)q_2/q_1}\}_{n\text{-even}}+\\
&\{\pm[j^{1-n+2r_1+2r_2}\cos((m-2r_2)\phi_{2t}+\\
&n-m-2r_1\mp1)\phi_{1t})]_{n-m-2r_1\mp1=(2r_2-}\\
&{}_{m)q_2/q_1}\pm[j^{1-n+2m+2r_1-2r_2}\cos((m-\\
&2r_2)\phi_{2t}-(n-m-2r_1\mp1)\times\\
&\phi_{1t})]_{n-m-2r_1\mp1=-(2r_2-m)q_2/q_1}\}_{n\text{-odd}})
\end{aligned}$$

$$\begin{aligned}
F_{q_2}=\sum&(a_n/2^{n-1})V_1^{\,n-m-1}V_2^{\,m}\,{}^{n}C_m{}^{n-m}C_{r_1}\times\\
&{}^{m}C_{r_2}(\{\pm[j^{-n+2r_1+2r_2}\sin((n-\\
&m-2r_1)\phi_{1t})+(m-2r_2\mp1)\times\\
&\phi_{2t})]_{n-m-2r_1=(2r_2-m\pm1)q_2/q_1}\pm\\
&[j^{-n-2m+2r_1-2r_2}\sin(-(n-m-\\
&2r_1)\phi_{1t}+(m-2r_2\mp1)\times\\
&\phi_{2t})]_{n-m-2r_1=(m-2r_2\mp1)q_2/q_1}\}_{n\text{-even}}+\\
&\{\pm[j^{1-n+2r_1+2r_2}\cos((n-\\
&m-2r_1)\phi_{1t}+(m-2r_2\mp1)\times\\
&\phi_{2t})]_{n-m-2r_1=(2r_2-m\pm1)q_2/q_1}\pm\\
&[j^{1-n+2m+2r_1-2r_2}\cos(-(n-m-\\
&2r_1)\phi_{1t}+(m-2r_2\mp1)\times\\
&\qquad\phi_{2t})]_{n-m-2r_1=(m-2r_2\mp1)q_2/q_1}\}_{n\text{-odd}})
\end{aligned}$$

The limits for the foregoing summations are

$n=1, 2, 3, \ldots$
$m=0, 1, 2, 3, \ldots$ up to n
$r_1=0, 1, 2, 3, \ldots$ up to $\leqq(n-m)/2$
$r_2=0, 1, 2, 3, \ldots$ up to $\leqq m/2$

Appendix II

The differential equation for the output voltage v of the oscillator in Fig. 11 is given by

$$d^2v/dt^2+(G/C)dv/dt+(1/C)di_p/dt+\omega_0^2v=0 \tag{21}$$

where i_p is the a-c component of the plate

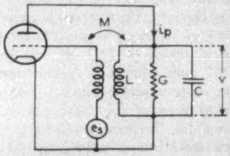


Fig. 11. Tuned-plate oscillator with a signal e_s applied in its grid circuit

current and $\omega_0^2=1/LC$. i_p may be represented as

$$i_p=\sum_{n=1}^{\infty}b_ne^n$$

and the composite voltage e may be represented as

$$e=[(1-\mu M/L)v+\mu e_s]$$

e_s is an external signal given by $e_s=E\sin\omega_st$ and μ is the amplification factor. The signal frequency $\omega_s/2\pi$ is near $(q_1/q_2)\omega_0/2\pi$, where q_1 and q_2 are simple integers. For a harmonic solution, the output voltage v may be assumed to be of the form

$$v=V_t\cos\int[(q_2/q_1)\omega_s+\phi_t']dt$$

Both V_t and ϕ_t' are slowly varying compared with the frequency of oscillation. Differentiating v once and twice and substituting in equation 21 gives an equation containing the following variables: ϕ_t', ϕ_t'', $\phi_t'^2$, V_t, V_t', V_t'', and

$$\begin{aligned}
&{}^{\sin}_{\cos}\int[(q_2/q_1)\omega_s+\phi_t']dt,\\
&\qquad{}^{\sin}_{\cos}\int m[(q_2/q_1)\omega_s+\phi_t']dt,\\
&\qquad{}^{\sin}_{\cos}\int n\omega_st,\ \text{ and }\ {}^{\sin}_{\cos}\int\{m[(q_2/q_1)\omega_s+\\
&\qquad\qquad\qquad\qquad\phi_t']\pm n\omega_s\}d_t
\end{aligned}$$

where m and n are integers.

As mentioned previously, the quasi-steady state method will be adopted; thus, V_t'', ϕ_t'', and $\phi_t'^2$ will be neglected. Again, V_t, V_t', and ϕ_t' could be satisfactorily considered constant for an interval containing several cycles of oscillations. For this small time interval, coefficients of

$$\cos \int [(q_2/q_1)\omega_s+\phi_t']dt \text{ and } \sin \int [(q_2/q_1)\omega_s+\phi_t']$$

could be separately equated with zero. By carrying out that process, two relations are obtained. Considering the first three terms of the plate-current series, the two relations for $q_1:q_2=1:2$ and 2:1, respectively, are

$$dV_t/dt=(1/2C)V_t\{[b_1(\mu M/L-1)-G]\times [1-(V_t^2/V_0^2)-2E_1^2/V_0^2]-(b_3/2)\times (1-\mu M/L)^3(E_1^2/V_t^2)\sin\phi_t\} \quad \text{(22A)}$$

$$(1/\omega_0)d\phi_t/dt=-Z_1+(1/4\omega_0C)(E_1^2/V_t)\times b_2(1-\mu M/L)^2\cos\phi_t \quad \text{(22B)}$$

$$dV_t/dt=(1/2C)V_t\{[b_1(\mu M/L-1)-G]\times [1-(V_t^2/V_0^2)-2E_1^2/V_0^2]+ b_2(1-\mu M/L)^2E_1\sin 2\phi_t\} \quad \text{(23A)}$$

$$(2/\omega_0)d\phi_t/dt=-Z_2-(2E_1/\omega_0C)b_2\times (1-\mu M/L)^2\cos 2\phi_t \quad \text{(23B)}$$

where

$$Z_1=(2\omega_s-\omega_0)/\omega_0,\ Z_2=2[(\omega_s/2)-\omega_0]/\omega_0,$$

$$E_1=\mu E/(\mu M/L-1), \text{ and}$$
$$V_0=\{[b_1(\mu M/L-1)-G]/(-3/4)b_3\times (\mu M/L-1)^3\}^{1/2}$$

Although equations 22(A), 22(B), 23(A), and 23(B) are valid for the small interval of the time constituting several cycles of oscillation, still they could be satisfactorily considered as the general instantaneous frequency-amplitude-phase relations of a triode oscillator when disturbed by an external signal; they represent the pulled oscillations of the system. As stated previously, the amplitude of oscillation V_t in the pulled state could be satisfactorily considered constant. By letting $V_t=V_1$, $\phi_t=\phi_{1t}$, and introducing a constant phase angle ϕ_2 in the applied signal, equation 22(B) reduces to equation 13. Also, letting $V_t=V_2$, $\phi_t=\phi_{2t}$, and introducing a constant phase angle ϕ_1 in the applied signal, equation 23(B) reduces to equation 14.

References

1. INTERNAL RESONANCE IN CIRCUITS CONTAINING NONLINEAR RESISTANCE, **R. Fontana.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 39, Aug. 1951, p. 945.

2. THEORETICAL AND EXPERIMENTAL INVESTIGATIONS OF THE BEHAVIOUR OF ELECTRON TUBE OSCILLATORS HAVING TWO DEGREES OF FREEDOM, **H. H. El-Sabbagh.** *M.Sc. Thesis*, Radio section, Electrical Department, Faculty of Engineering, Alexandria University, Alexandria, Egypt, June 1953. (Prepared under author's supervision.)

3. SIMULTANEOUS OSCILLATIONS IN OSCILLATORS, **J. S. Schaffner.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., no. 2, June 1954, p. 2.

4. A STUDY OF LOCKING PHENOMENA IN OSCILLATORS, **R. Adler.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 34, June 1946, pp. 351–57.

5. THE FREQUENCY SPECTRUM OF A PULLED OSCILLATOR, **T. J. Buchanan.** *Ibid.*, vol. 40, Aug. 1952, p. 958.

6. ANALYSIS OF A TRIODE OSCILLATOR WITH LOSSES ASSOCIATED WITH THE INDUCTIVE BRANCH AND SIGNAL APPLIED IN GRID CIRCUIT, **Abd El-Samie Mostafa.** *AIEE Transactions*, vol. 74, pt. I, 1955 (Jan. 1956 section), pp. 777–81.

7. A NEW METHOD FOR TREATING ELECTRON TUBES WHEN USED AS SUPERREGENERATIVE DETECTORS—PART III. EXPERIMENTAL INVESTIGATION OF SUPERREGENERATIVE CIRCUITS, **Abd El-Samie Mostafa, M. El-Shishini.** *Ibid.*, vol. 72, pt. I, July 1953, pp. 290–97.

8. VARIABLE ELECTRIC CIRCUIT THEORY WITH APPLICATIONS TO THE THEORY OF FREQUENCY MODULATION, **J. R. Carson, T. C. Fry.** *Bell System Technical Journal*, New York, N. Y., vol. 16, Oct. 1937, p. 513.

Mass Flowmeter with Retarding Disk Integration

R. G. BALLARD
MEMBER AIEE

J. R. MACINTYRE
FELLOW AIEE

THE FIRST large-scale use of flowmeters based on the conservation of momentum principle was employed in the B-47 airplane to measure fuel flow rates to the jet engines. This paper describes a flowmeter for gases which, although operating on the same basic principle, uses a retarding disk and register to integrate the flow rate. This flowmeter will directly measure the mass, or standard cubic feet, of gas consumed. It is much smaller and lighter in weight than conventional meters used for this purpose and may prove to be a most important contribution to the gas measuring industry. The basic principles of gas measurement have remained unchanged since 1844. Progress in the art has been largely confined to improved materials. It is the purpose of this paper to report on the preliminary development of the flowmeter, to discuss principles of operation, and to review test results obtained on the experimental sample.

The conservation of momentum principle of flow measurement may be described by reference to Fig. 1. An impeller with longitudinal peripheral slots, rotated at constant speed, imparts angular momentum to the fluid. The total momentum is the product of mass flow rate, angular rotational speed, and a suitable constant. A turbine, similar in mechanical construction to the impeller, is restrained in its angular motion by a retarding disk, a spiral spring, or other means. This turbine removes the angular momentum previously imparted by the rotating impeller. The torque produced, and hence the rotational speed, or angular deflection, of the turbine is proportional to mass flow rate.

Construction

Specific application of this principle to the measurement of gas required integrated flow measurements rather than rate of flow. One of the simplest integration methods appeared to be a temperature-compensated combination of a retarding disk and retarding magnets substituted for the restraining spring used in the indicating system. The total revolutions of the disk are then proportional to the integrated flow. This introduces certain problems which will be discussed later.

A sample of a flowmeter constructed in this manner is shown in Fig. 2. The retarding disk and magnet assembly, and the revolution-counting register are clearly shown. The impeller and turbine are located in the lower portion of the meter, and are completely enclosed by the casing.

The mechanical construction of the synchronous motor drive for the impeller is of particular note. The hysteresis material for the rotor forms the outer shell for the impeller. The stator is of a closed construction thereby permitting the inside surface of the laminations to be the seal which prevents the gas from entering the stator windings. A maximum operating efficiency of 45% (per cent) was attained with this design. The total power input was 3 watts at 120 volts 60 cycles.

The register for indicating flow is similar to a watt-hour meter register and is driven by a worm gear on the turbine shaft. One revolution of the turbine shaft represents approximately 0.08 pound (approximately $1^2/_3$ standard

---

Paper 57-113, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 24, 1956; made available for printing December 11, 1956.

R. G. BALLARD and J. R. MACINTYRE are with the General Electric Company, West Lynn, Mass.

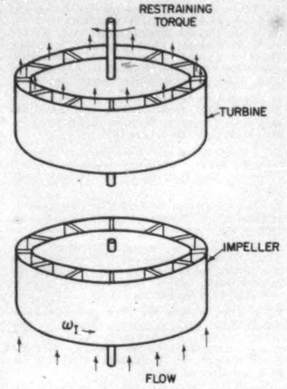


Fig. 1. Schematic sketch illustrating principle of operation of mass flowmeter

cubic feet of natural gas) and the gear reduction provides one unit indication for each 22 pounds (500 cubic feet) of gas metered. These constants, of course, may be readily changed to suit requirements.

Test Methods and Results

Calibration of this meter, at the higher flow rates, has been accomplished by the use of a low-pressure, orifice-plate type of flow prover, and at the lower flow rates by a diaphragm-type positive displacement meter. For convenience, air was used as the test gas.

Air flow is maintained through the meter and the orifice by means of a motor-operated blower. It was found convenient to control the flow-rate by the use of an adjustable autotransformer supplying power to the blower motor. The retarding disk was visible during these tests so that the revolutions per minute could be checked with a stop-watch. Accuracy of individual checks is estimated to be approximately one half of one per cent.

The meter under test is shown in Fig. 3. With a flow rating of 94 pounds (1,250 cubic feet) per hour of air at atmospheric pressure, the turbine speed was approximately $1\frac{1}{2}\%$ of the impeller speed. This rating is based upon a pressure drop across the meter of $\frac{1}{2}$-inch water column.

Calibration is shown in Fig. 4 after adjustment for best linearity over the range of 15% to 100% rating. The departure from linearity below 15% is caused by coupling between the impeller and turbine and by friction. It should be noted that the meter was still linear within 1.0% up to the limit of the test equipment at 340 pounds (4,500 cubic feet) of air per hour (350% of rating). In applications where the additional pressure drop at this high flow can be tolerated, the meter could be considered to have the 340-pound rating and be accurate within 1.0% down to approximately 4% of rating.

Temperature error in the meter centers principally in variation of retarding torque due to temperature coefficients of the aluminum disk and the magnets. Compensation for these errors is achieved by mounting the magnets on bimetal supports so that the magnet gap is varied with temperature changes. In the sample, the compensated temperature error was -0.3% per degree centigrade over the temperature range of -10 to $+50$ centigrade based on a nominal temperature of 30 centigrade. This could be improved with minor adjustments.

Discussion

Linearity

To obtain correct integration, it is essential that the rotational speed of the turbine be exactly proportional to the mass flow rate. This condition does not exist in the normal retarding disk case for the following reasons.

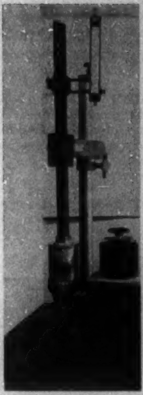


Fig. 3. Test setup of flowmeter and flowprover

Using parameters from the experimental sample, the impeller rotates at a constant speed (ω_I) of 100%, and sufficient restraint is provided by the retarding disk so that the turbine speed (ω_T) is 1.56% of the impeller speed at 100% flow.

Since the turbine is turning at 1.56% of the impeller speed, it is not removing all of the angular momentum in the flowing gas, and 1.56% of the momentum is still in the gas stream as it leaves the impeller The expression for the turbine speed under these conditions can be written as

$$\omega_T = \frac{0.01585}{1+.0001585\dot{M}}\dot{M} \qquad (1)$$

where

ω_T = speed of the turbine, per cent of the impeller speed
$\dot{M}$ = mass flow, per cent of nominal rating

This is not a linear curve because of the 0.0001585 $\dot{M}$ term in the denominator. Let it be assumed that the desired linear relationship ω_{T_L} is that the produced by the initial slope of the ω_T equation or

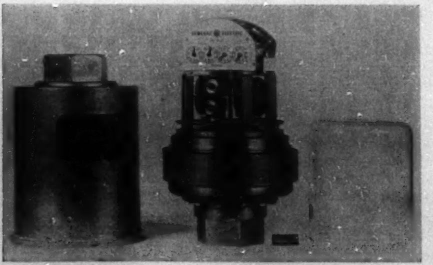


Fig. 2. Model of retarding disk mass flowmeter

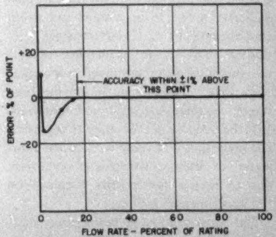


Fig. 4. Calibration curve of mass flowmeter

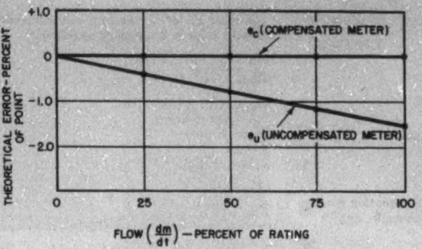


Fig. 5 (above). Theoretical calibration curves for compensated and uncompensated mass flowmeters

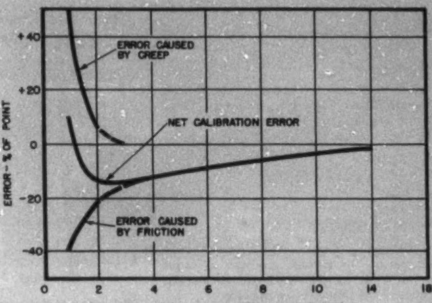


Fig. 7 (right). Mass flowmeter coupling and friction errors

$$\omega_{T_L}=0.01585\dot{M} \tag{2}$$

where ω_{T_L} = the turbine linear speed. The per cent of point error e_u of the actual uncompensated turbine speed from the desired speed is derived as shown in the Appendix and is expressed as

$$e_u=\frac{0.01585\dot{M}\%}{1+0.0001585\dot{M}} \tag{3}$$

Fig. 5 shows a plot of this theoretical uncompensated meter error versus the flow rate.

The correct linear relationship can be attained by adding speed to the turbine to compensate for the loss in speed caused by the incomplete removal of the angular momentum. Such a factor was supplied by bending the outlet ends of the turbine, as indicated in Fig. 6, to add a speed. Speed added in this manner can be expressed as

$$\Delta\omega_T=0.00025\dot{M}V\% \tag{4}$$

where

$\Delta\omega_T$ = added speed, per cent of impeller speed
$\dot{M}$ = mass flow, per cent of nominal rating
V = factor proportional to velocity of flow (unity at 100% flow, 0.5 at 50%, etc.)

The complete theoretical compensated turbine speed ω_{T_C} may now be written as

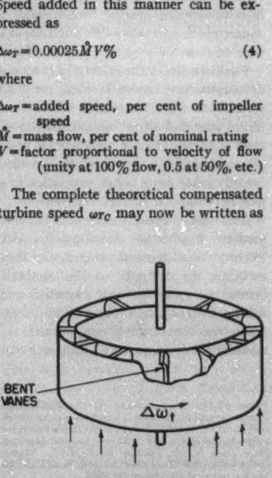


Fig. 6. Compensation scheme used on mass flowmeter

$$\omega_{T_C}=\omega_T+\Delta\omega_T=\frac{0.01585}{1+0.0001585\dot{M}}\dot{M}+0.00025\dot{M}V \tag{5}$$

where ω_{T_C} = compensated turbine speed, per cent of impeller speed.

The per cent of point deviation from linear, or theoretical, error may be written as the following

$$e_c=\frac{\omega_{T_L}-\omega_{T_C}}{\omega_{T_L}}100\%$$

$$=\frac{0.01585\dot{M}-\left[\frac{0.01585\dot{M}}{1+0.0001585\dot{M}}+0.00025\dot{M}V\right]}{0.01585\dot{M}}100\% \tag{6}$$

This is plotted in Fig. 5. The error is too small to be apparent, and is less than a maximum of two or three hundredths of one per cent.

The compensated meter has a density error as indicated by the ρ term included in $\dot{M}$ in the compensation equation $\Delta\omega_T=0.00025\dot{M}V$. However, since $\Delta\omega_T$ in the case of the meter under discussion is approximately 1.5%, variations in ρ of 10% would produce errors in $\Delta\omega_T$ of only 0.15%.

INACCURACIES AT LOW FLOW RATES

The combination of friction, and coupling between the rotating impeller and retarding disk, causes the inaccuracy at the lower flow rates shown in Fig. 4. The coupling is caused by the viscous drag between the impeller and turbine and the circulation of gas between them because of the action of centrifugal forces on the gas particles. This effect becomes negligible at flows above 2% of rating. A complete quantitative analysis of the friction and coupling factors has not been made, but Fig. 7 illustrates the cause of the peculiar shape of the calibration curve at low flow rates. As may be seen, the shape of the calibration curve below 15% is caused by the algebraic addition of the creep curve and the friction curve. It may be seen further that the magnitude of the creep torque at low flows exceeds the magnitude of the friction torque, and therefore the retarding disk will rotate at conditions of zero flow and produce erroneous integration readings. A magnetic brake may be used to stop the meter at low flow rates, but this type of solution obviously makes for large calibration errors at low flow rates as the braking point is approached. At present, this creep at zero flow is a serious limitation to certain applications for the meter. However, it is anticipated that further development will materially reduce or eliminate this difficulty.

Conclusions

The practicability of measuring mass rate of flow of gases, and providing direct integration, has been demonstrated. Problems of possible errors with variation in gas density, and with temperature changes, have been discussed. In addition, a discussion of creep at zero flow and nonlinearity at low flows has been presented.

In conclusion, it should be noted that the density problem can be eliminated and the creep and friction errors reduced by using another system based on the same mass flow principle. This system uses a spring restrained turbine and a servo-operated mechanical integrator. It is proposed that this system will be the subject of another paper.

Appendix

The torque produced by the retarding disk is proportional to the angular velocity of the turbine attached to it.

$T_D = K_1 \omega_T$

T_D = torque on disk

ω_T = turbine speed

K_1 = a constant

The angular momentum of the flowing gas and the available torque is proportional to the angular velocity of the impeller and the mass flow rate.

$T_I = K_2 \omega_I \mathring{M}$

T_I = torque available in gas

ω_I = impeller speed

$\mathring{M}$ = mass flow rate

K_2 = a constant

The angular momentum removed from the gas by the turbine produces a torque proportional to the difference in angular speeds of the impeller and turbine.

$T = K_3(\omega_I - \omega_T)\mathring{M}$

T = torque from angular momentum removed from gas

At any equilibrium condition, the torque on the disk must equal the torque produced by removal of angular momentum from the gas.

$T_D = T = K_1 \omega_T = K_3(\omega_I - \omega_T)\mathring{M}$

K_4 = a constant

From tests on the experimental sample

at $\mathring{M} = 100\%$

$\omega_I = 100\%$ impeller speed

$\omega_T = 1.56\%$ impeller speed

$1.56 = K_4(100 - 1.56)(100)$

$K_4 = 0.0001585$

$\omega_T = 0.0001585(100 - \omega_T)\mathring{M}$

$$= \frac{0.01585}{1 + 0.0001585\mathring{M}}\mathring{M} \qquad (7)$$

The initial slope of this equation is taken as the desired linear relationship, i.e.,

$$\omega_{TL} = 0.01585\mathring{M} \qquad (8)$$

ω_{TL} = desired linear speed, per cent of impeller speed

The per-cent deviation from linear, or per-cent error, may then be expressed as e_u where

$$e_u = \frac{\omega_{TL} - \omega_T}{\omega_{TL}} 100\%$$

substituting the equations given above for ω_{TL} and ω_T

$$e_u = \frac{0.01585\mathring{M}\%}{1 + 0.0001585\mathring{M}} \qquad (9)$$

The speed added by the bent vanes to correct the error is determined as follows:

$\Delta\omega_T = \omega_{TL} - \omega_T = K_5 \rho V^2$

where

$\Delta\omega_T$ = speed to be added, per cent of impeller speed

ρ = gas density

V = gas flow velocity

K_5 = a constant

$\rho V = K_6 \mathring{M}$

$\Delta\omega_T = K_7 \mathring{M} V$

at 100% flow, then

$V = 1.0$ (0.5 at 50%, etc.)

$\omega_{TL} = 1.585$

$\omega_T = 1.560$

$K_7 = 0.00025$

and $$\Delta\omega_T = 0.00025\mathring{M}V\% \qquad (10)$$

References

1. The Momentum Principle Measures Mass Flow Rate, V. A. Orlando, F. B. Jennings. *ASME Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 76, no. 6, Aug., 1954, pp. 961–65.

2. Displacement Gas Meters. *Handbook E-4*, American Meter Company, New York, N. Y., 1956, p. 6.

Buried Cable Telephone Systems

G. L. CHILBERG
NONMEMBER AIEE

HISTORICALLY, both the power and telephone operating companies throughout the country have provided distribution facilities into residential areas by means of aerial construction. This has followed logically from the fact that aerial construction has been the most economical type of facility available to provide such services.

Of late, however, certain developments have taken place which induced the Bell Telephone System to look critically at conventional aerial distribution systems. The question has been raised as to the feasibility of providing economical buried distribution systems in new residential developments.

Often in certain of the more exclusive developments all utilities are provided on an underground basis at the request of and paid for by the owner or developer. However, such "tailor-made" buried distribution systems have generally been at such a cost as to discourage their use except in exceedingly exclusive and demanding situations. The question coming to the fore of late has been, "Could not the excess cost of providing such buried facilities be reduced substantially, if not eliminated, under average circumstances?"

Buried Telephone Distribution

Like all other utilities, the telephone industry has just one commodity to sell, namely, service. To meet that responsibility, however, the industry must provide not just service as such. It must provide the kind of service the customer wants in the manner the customer wants and with the maximum reliability that is economically practical. In addition, that service should obviously be engineered in such a manner as to reduce to the practical minimum the hazards of placement and maintenance to which the construction personnel are exposed.

APPEARANCE

In recent years, there has been a drastic change in the type of and manner of providing new individual residence units, particularly in the faster growing areas of America. Prior to World War II, a substantial part, if not the majority, of residential development was built individually on lots located at random as selected by the owner. In addition, a rather large percentage of these individually built houses were 2-story in design or, if 1-story, were of relatively high ridge-pole construction.

The modern trend, however, is toward mass development with a large proportion of ranch-house or low-eave design. In such circumstances, the houses are built in developments ranging from possibly a square block up to literally thousands at a time on essentially an assembly line basis.

Furthermore, in the interest of economy of construction, most builders strip all trees from the construction area as an initial step and either replant with small trees or leave it to the house owner to do all landscaping. Consequently, it frequently takes years before vegetation in a new development reaches the rooftops.

As a result, aerial utility structures in modern residential developments are likely to be all too conspicuous, and they certainly do not add to the aesthetic beauty in spite of their essential and utilitarian functions. Figs. 1 and 2 illustrate the contrasting appearances of residential developments with conven-

---

Paper 57-140, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted November 8, 1956; made available for printing December 10, 1956.

G. L. Chilberg is with the American Telephone and Telegraph Company, New York, N. Y.

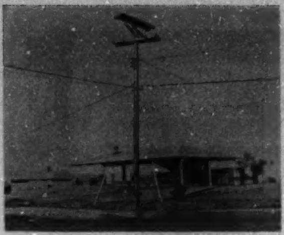


Fig. 1. Typical subdivision with conventional aerial distribution

Fig. 2. Typical subdivision with buried distribution

tional aerial as contrasted to buried distribution systems respectively.

RELIABILITY

Those locations in the country which are in the area of the sleet belt or have been subject in recent years to the all too frequent hurricanes, cannot help but be conscious of the communication impairments which have followed such disasters. Analyses conducted after several of the most recent and severe of these storms, indicated that at least 90% (per cent) of the telephone service impairments were the direct result of the physical exposure inherent in aerial plant. Despite all the reasons that service impairments may be deemed inevitable, the customer's attitude is, understandably, that he is entitled to service at all times and under all conditions, regardless of such catastrophies as hurricanes or inclement weather. The industry can not question the customer's right to such expectations. However, it is not economically feasible at rates which customers would be willing to pay to construct aerial plant that is completely invulnerable to the most severe storms. Buried plant, on the other hand, is safe from all weather conditions except, perhaps, severe floods.

SAFETY

All industry has a practical as well as a moral obligation to engineer its facilities so as to reduce to a practical minimum the accident hazards to which its workmen are subjected. During the past 10 years, 39.7% of all lost-time injuries suffered by plant men in the Bell System have been a direct or indirect result of working aloft. This is in contrast to the very much smaller per cent of craft man-hours spent aloft.

Obviously, to the extent that the need for working aloft can be eliminated, the industry will provide a safer place for employees to work with a comparable reduction in suffering and anguish by them and their families. The foregoing factors certainly justify a firm objective of placing plant underground to the maximum extent and as rapidly as it is economically feasible.

Developments Facilitating Buried Distribution

Prior to World War II, the conventional facility for providing large numbers of physical telephone circuits consisted of lead sheath cable containing paper or pulp insulated conductors. This type of cable was quite satisfactory for aerial construction, although its weight, due to the lead sheath, required very substantial and, consequently, relatively costly supporting structures. In the underground, however, the vulnerability of the lead sheath to various types of corrosion when in immediate contact with the soil required either conduit structures to house it or asphalt, paper, and jute protection on the outside of the sheath to protect against soil corrosion. In addition, the vulnerability of the paper or pulp conductor insulation to even slight amounts of moisture necessitated hermetically sealed terminals which were of necessity expensive to produce and install. Furthermore, the sealed chamber necessary in these terminals to ensure their absolute moisture tightness made any subsequent rearrangements a time-consuming process.

PLASTIC-SHEATH CABLE

Extensive research work had been conducted prior to World War II, aimed toward the development of various types of plastic. During the war, many of these plastics and, of particular interest to the Bell System, polyethylene, came into quantity production.

Immediately following the war, the tremendously expanded demand for telephone cable would have required more lead for sheathing than was physically available on the world markets. Accordingly, the Bell System turned to plastics as a substitute. The result was the extensive use for cable sheathing of polyethylene in lieu of lead, with an inner shield of aluminum, in some instances reinforced with a second inner shield consisting of a soldered, corrugated steel tube to ensure greater physical protection and moisture imperviousness.

This development had as a by-product certain characteristics which made the direct burial of telephone cable far more feasible. Plastic-sheathed cable, commonly known in the Bell System as "alpeth" when the sheath consisted of polyethylene with an inner shield of aluminum, or "stalpeth" when the sheath consisted of polyethylene on the outside with inner shields of aluminum and steel tape, proved to be substantially less costly and relatively invulnerable to corrosion when in direct contact with the earth. In addition, it was lighter and much easier for the craftsman to handle. A very important added advantage was that the use of polyethylene rather than lead for the outer sheath eliminated the high extrusion temperatures associated with lead cable and, hence, paved the way for plastic insulation on the individual conductors.

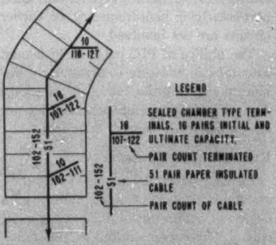


Fig. 3. Typical distribution cable for residential block-paper insulated cable with sealed-chamber terminals

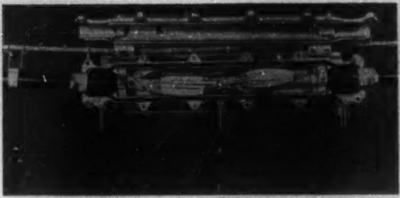


Fig. 4. Typical sealed-chamber terminal showing internal splicing of cable conductors to conductors from face plate binding posts

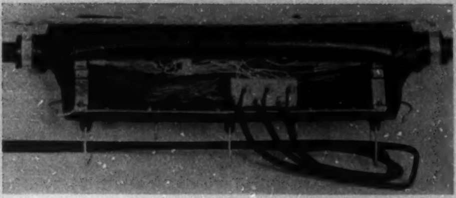


Fig. 5. Typical ready-access terminal for aerial use showing complete availability of all cable conductors

Plastic-Insulated Conductors: PIC Cable

As previously stated, conventional telephone cable, up to the end of World War II, consisted of lead sheath with paper- or pulp-insulated conductors. The paper or pulp insulation on the individual conductors was the most economical and practical at the time. This was because satisfactory plastic for such insulation was not available in commercial quantities.

The advent of plastic sheath cable, however, and the availability in commercial quantities of plastics with the necessary physical characteristics, notably polyethylene, paved the way for an entirely different type of cable, namely, polyethylene-insulated conductor cable, commonly known as PIC cable.

This type of cable, now in large-scale quantity production, gives promise of an entirely new type of facility for telephone distribution with potentialities that may well revolutionize conventional forms of telephone distribution systems. In its standard forms, it has a wire-to-wire and core-to-sheath dielectric strength of up to 10,000 volts, depending upon the wire gauge (conductor insulation thickness varies with the conductor gauge). It is completely impervious to atmospheric moisture and can even stand a certain amount of physical water in the core, particularly if high-frequency or carrier circuits are not involved.

The ability of PIC cable to withstand atmospheric moisture is particularly significant as facilitating buried installations. Also, it permits application of what are known as "ready-access" terminals.

Ready-Access Terminals

An economical telephone distribution system is dependent upon a very high degree of utilization of the available circuits. This is complicated by the transitory and changing demands of the public in the territory served. There is a high turnover of demand for lines as people move with the consequent disconnections and reconnections, usually in different locations. Party-line service, both 2-party and 4-party, and in some instances with a choice between message and flat-rate service, complicates the problem of realizing the optimum load on a given number of circuits. A pair of wires on a 2-party line is only 50% efficient when serving one customer. A pair of wires on a 4-party line is only 75% efficient if it serves only three customers. Accordingly, the terminal appearances of individual pairs in a sealed terminal chamber distribution system must be such as to provide a selection of customers adequate to permit an optimum fill for all working pairs.

The foregoing necessitates a complicated multipling arrangement with conventional paper-insulated cable to facilitate maximum utilization. An example of this multipling arrangement is shown in Fig. 3. These multipling arrangements are engineered on the best information available prior to construction. However, variations in growth frequently necessitate changing the count appearing in a given terminal at a very considerable amount of engineering and construction expense. This is time-consuming and difficult if the distribution cable is paper-insulated with sealed-chamber terminals, particularly if the system is buried. A typical terminal used in such cases to make the cable conductor pairs available for the service wires is shown in Fig. 4.

The PIC cable, being able to withstand atmospheric moisture within the core, permits an entirely new principle in cable distribution. This is through the use of the so-called ready-access terminal shown in Fig. 5 for aerial distribution and, in Fig. 6, for buried distribution systems. It will be noted that the entire cross section of the cable is exposed at each terminal point, thereby permitting complete flexibility and complete availability of every pair in the cable without expensive rearrangement work.

A typical multipling layout for PIC cable using the foregoing ready-access terminals is shown in Fig. 7. With this facility, it is at last feasible to provide a buried distribution system without incurring the possibility of inordinate rearrangement expense due to changes in cable-multipling arrangements if unexpected variations in the type or amount of customer demand should develop.

Placing Equipment

For many years, the Bell System has found it feasible to bury long-distance cables directly in the ground at a cost comparable with or below that of aerial plant. This was done by the use of heavy plowing equipment, an example of which is shown in Fig. 8. The economic application of this heavy equipment, however, necessitated large jobs and open terrain where uninterrupted plowing was feasible.

Another type of development which can be used in less populated urban areas, and has proved particularly suitable in rural areas, is a light plow suitably modified to permit placing the cable or service wire

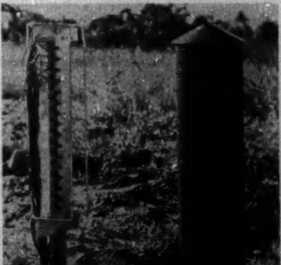


Fig. 6. Typical ready-access terminal of the pedestal type showing complete availability of all cable conductors

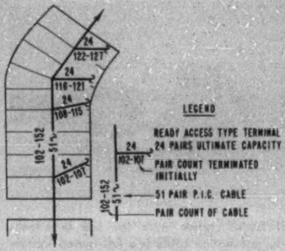


Fig. 7. Typical distribution cable for residential block: PIC cable with ready-access type of terminals

directly through the share into the bottom of the cut. One example of this type of unit is shown in Fig. 9. It is usually operated by a 2-man crew and, in favorable soil, can place a service wire or a small cable at a depth of 18 to 20 inches and at a rate of better than one mile an hour.

In typical city subdivisions, however, we have not yet found it practical to place buried distribution plant by the plow method. Efforts are continuing to develop techniques and equipment which will permit the plowing-in of such plant because of the greater potential economy of the plowing method. However, to date, trenching has been found to be the most practical method of placing distribution plant, and particularly service wires, in urban developments.

Certain manufacturers have recently developed small trenching machines which, with further refinement, give promise of trenching up to the foundation of the house on an economical basis and with a minimum of equipment investment per unit. An example of this type of equipment in its present form is shown in Fig. 10. This unit digs a trench either 3 or 5 inches wide and as much as 40 inches deep, depending upon the bucket line used. In favorable soils, it can dig up to 15 or more feet a minute. It is operated by one man and can be towed behind a passenger car or pick-up truck at about 25 miles an hour with one man conducting the entire operation including coupling and uncoupling from the tow vehicle.

There is every indication that the increasing demand for these types of placing equipment will stimulate industry to refine existing units and devise improved units to the end that buried facilities may be placed more economically and with a lesser investment in heavy equipment.

Current Buried Distribution Systems

As of today, trial buried telephone distribution systems are being installed in most of the Bell System operating companies. In many of these trial installations, buried plant is being placed in accordance with these modern economical methods in one section of a subdivision with conventional aerial distribution placed in another section of the same subdivision, both parts on a "keep cost" basis, to permit cost comparisons.

Generally, the buried telephone distribution systems fall into two basic categories. These are: 1. "full buried" systems with a cross-connection box at the entrance to each block or group of blocks to provide necessary flexibility, or 2. "pedestal terminal" systems where the flexibility is provided by connecting the buried service wires from the residences to pedestal terminals mounted along the rear property line, usually one for about every four lots. In the full buried systems, the service wires to the customer's premises are connected to the appropriate pairs in the distribution cable through a buried splice case such as is shown in Fig. 11. In the pedestal terminal systems, the service wires are connected to the appropriate cable pairs at a terminal, such as is shown in Fig. 6, nearest to the premises.

In either of these systems, it has been found most practical in our experience thus far to place the telephone distribution cable along a rear property line easement which can be shared jointly with power. When this is done, it has been found desirable to place the telephone distribution cable 2 to 3 feet to one side of the rear property line at a depth of 24 to 36 inches, with power a similar distance the other side of the rear property line and at a depth of 36 to 48 inches. The telephone service wires are then run at the same depth as the telephone distribution cable from the buried splice case or the pedestal terminal either directly to the served residence or along the trench with the distribution cable to a point nearest the residence and then to the house. The choice of method depends upon the relative location of the house and terminal and the cost factors relating to the cost of wire versus the cost of trenching for the service wire.

To date, the width of trench has been controlled almost entirely by practical design limitations in the available trenching equipment. A minimum width of trench is most satisfactory because of the lesser cost of backfilling and the elimination of tamping problems. So far, a trench of approximately a 3-inch width has proved most satisfactory, and it can be produced by at least two of the trenching machines in commercial production.

RELATIVE COST: BURIED VERSUS AERIAL DISTRIBUTION SYSTEMS

The costs of both aerial and buried distribution systems are influenced to a major degree by conditions at each location. For example, aerial distribution costs are adversely affected by the excessive number of anchors and guys required in the winding property lines now gaining prominency in many subdivisions. Where local conditions dictate street rather than rear property line distribution,

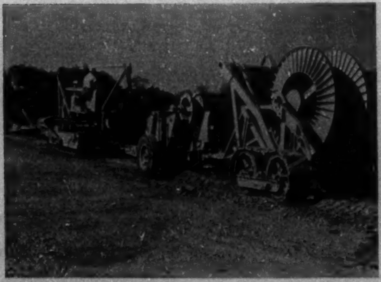


Fig. 8 (left). Typical plow train for direct burial of long-distance cables

Fig. 9 (right). Typical tractor and attached plow unit for direct burial of small cables or service wires

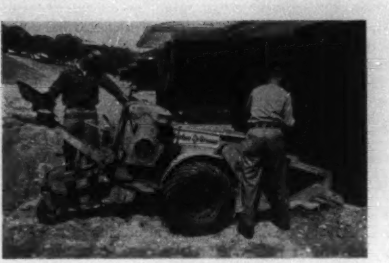


Fig. 10. Small trencher now available for use in urban buried distribution work

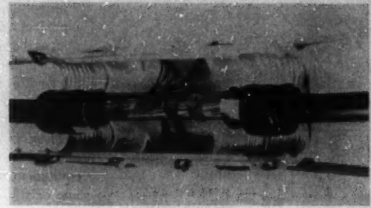


Fig. 11. Small buried splice case for use in attaching service wires to distribution cable in a full-buried system

excessive costs for aerial plant are likely to result from the need for high poles to provide drop clearances to the houses on the opposite side of the street, emphasized of late by the low roof line used in many housing developments. If these cost penalties are not reflected in the form of higher poles, they are suffered later by the installation forces in gaining clearance across the street by other means.

Buried distribution costs, on the other hand, are adversely affected by: 1. unusually hard or rocky soil, 2. building techniques by subdividers wherein topsoil is piled along the rear property line, thus impeding the efficient programming of the buried distribution job, 3. in northern climates during periods when the ground is frozen to a depth of more than a few inches, and other special conditions. There are also situations where residential developments are built on essentially rock ledge formations with a minimum, if any, topsoil. In such cases blasting is frequently required to place such necessary underground facilities as sewer, water, and gas.

However, actual cost results realized to date in various trial installations, indicate that for new subdivisions, and if digging conditions are favorable, buried telephone distribution systems can be placed by today's equipments and techniques at costs generally favorable as compared to costs of conventional aerial construction. This is particularly true where the subdivision is laid out with winding streets necessitating the large number of anchors and guys thereby required for aerial plant. However, the practicability of economical buried telephone distribution placement is sorely handicapped unless the development is fully built up on every lot at the time of initial construction thereby permitting complete placement of the buried distribution system prior to landscaping by the eventual owners.

Principal Remaining Problems

The Bell System is still some distance from being able to provide buried distribution systems without cost penalties and, hence, without special charges to the customer or developer. The principal problems remaining before this can be accomplished are described in the following paragraphs.

It is necessary to have a conductor joint for PIC cable at approximately the same cost as is presently entailed in the conventional "pigtail" conductor joint used on paper or pulp cable, but providing the same hermetic integrity as the polyethylene-insulated conductor itself. Developmental work is in progress on such a joint and the present status promises early success. In the intervening period, however, it is necessary to insert the pigtail conductor joint in a special silicone-filled sleeve to give moisture imperviousness. This method is too expensive if applied to a cable splice having a large number of pairs.

Where these PIC cables are spliced to conventional cables with paper or pulp-insulated conductors, it is necessary to "plug" the PIC cable at the splicing point to prevent moisture from traveling from the impervious PIC to the conventional paper-insulated conductor cable. This can be done at present but the cost of these plugs is still too high to permit widespread installations. Work is progressing on the development of a more inexpensive plug. The present status of this project suggests an early successful solution.

Mention was made of the problems attendant upon buried distribution systems in developments on so-called rock ledge formations. Developmental work is presently progressing aimed toward techniques which will permit placing the telephone distribution system in the same trench which must necessarily be provided for water, sewer, and gas. The practical problems of a solution such as this are fully recognized, but its advantages from the viewpoint of economical power and telephone buried-distribution installations, where soil conditions are severe, are such that it cannot be ignored.

Where telephone and power share a common easement along the rear property line, there is the problem of joint power-telephone co-ordination with reference to horizontal and vertical clearances. This problem is being investigated by the Edison Electric Institute–Bell System Joint Committee on Plant Co-ordination which is now reviewing common problems in connection with buried distribution systems.

None of the present trenching devices is entirely satisfactory. Several manufacturers of commercial equipment are attacking this problem. Some of the devices presently available give promise of meeting the requirement, and it is expected that the problem will be overcome in the relatively near future.

In a full buried distribution system, the service connection wires from the residence to the distribution cable must be connected into that cable at points relatively close to each residence. Small mechanical splice cases (Fig. 11) are available for this purpose. However, their cost makes them unattractive. Development work presently in progress gives reason to expect that an inexpensive plastic-sheath closure will soon be available to permit splicing the service wires directly into the distribution cable, and closing it with a minimum of manpower and costs.

As previously stated, PIC cable can withstand any amount of atmospheric moisture within the core. However, in the event of a sheath damage which would permit the entrance of physical water, there is a requirement for certain restoration devices or techniques which are not yet fully resolved.

1. Inasmuch as evidence of water in the core may be some distance from the point of sheath damage where the water entered the core, it has been difficult in the past to locate the point of sheath damage. What appears to be an effective and dependable method for locating such buried-sheath damage points has now been developed and confirmed by preliminary field trials. The final development, refinement, and supply of this method and the associate devices, will be completed in the immediate future.

2. With paper- or pulp-insulated cable, wet spots can be easily and positively located by "burning in" the trouble with high-voltage battery. This provides a solid short circuit at the previous high-resistance "wet short" which can be positively and easily located by tone and locating coil. With PIC cable, wet shorts cannot be "burned solid" with high-voltage battery and, hence, are difficult to locate. This problem is the subject of extensive research effort, and results to date give reason to expect the development of an early successful method.

3. In paper- or pulp-insulated cable, moisture in the core causes the paper to swell, thus providing a block which prevents the water from traveling more than a few feet from the point of entry through the sheath. The conductor insulation in PIC cable, on the other hand, does not swell when exposed to moisture. Hence, water entering a PIC cable through a sheath defect may travel for hundreds of feet in both directions from the point of entry. The problem hence arises as to the most economical and effective way to expel this moisture after the sheath defect is located and repaired. Several possible solutions to this problem are being investigated but, to date, there has not been developed a fully satisfactory answer.

Conclusions

In the approximately 2 years since the Bell System started aggressive research and developmental efforts aimed toward buried distribution systems, much has been accomplished. Trials are presently in progress in most of the states of the Union using the new methods and materials recently developed. There is every indication that in the near future it may be possible to provide telephone buried distribution systems at negligible, if any, additional cost where and when conditions are favorable for buried plant. However, in order to provide reasonable application of the buried distribution philosophy in new developments, it will be necessary to produce substantial further economies. Not only must there be the ability to complete the average installation without cost penalty, but the economies in favorable installations must permit absorbing the penalties in unfavorable locations and during periods which are unfavorable for burying. The Bell System has every expectation that this can and will be accomplished in the foreseeable future.

Proposed Size Standards for Toroidal Magnetic Tape-Wound Cores

AIEE COMMITTEE REPORT

THE MATERIALS Subcommittee of the AIEE Magnetic Amplifier Committee was organized principally to set up standards and to encourage the interchange of information on materials used in conjunction with magnetic amplifiers. One of the projects undertaken by the Subcommittee has been the standardization of magnetic tape-wound toroidal-core sizes. To accomplish this objective, a separate Working Group was formed in Nov. 1953 and charged with the responsibility of developing a list of recommended standard nominal sizes for tape-wound toroidal cores. The proposed standards in this report were established by the working group in June 1956, approved by the Materials Subcommittee in July 1956, and adopted by the AIEE Magnetic Amplifier Committee in Nov. 1956.

This is the second report presented on the activities of the Materials Subcommittee. The first report[1] set forth standards for the presentation of basic information on core magnetic materials used in magnetic amplifiers. Primarily, because of the very rapid growth in the field during the past few years, it has been most difficult to set standards on cores or core materials used in magnetic amplifiers. Design and manufacturing techniques were to a great extent of an empirical nature, and the level of production was restricted so that trends were difficult to foresee. During these years, however, the base of theoretical knowledge in the field has been greatly strengthened and there has been a widespread increase in the level of production. Consequently, it may be anticipated that additional reports on recommended standards in the field will be forthcoming. In particular, it is expected that a report on recommended standard methods for core grading and matching will be presented in the near future.

Objectives

Those engaged in the manufacture or use of toroidal cores will both recognize the necessity for a size standardization, although for somewhat different reasons. The designer who is called upon to develop some form of magnetic circuitry, using toroidal cores, would prefer to have a wide choice of sizes available from which to choose. However, if an unlimited range exists, each size must itself be special, and as such will not be available as readily from the manufacturer. Since each core size would be made in lesser volume, unit costs would tend to be higher on all sizes.

The manufacturer will benefit from size standardization in several ways, not the least of which is the ability to maintain stock supplies for rapid delivery. Effective manufacturing controls can best be maintained by continuous production of a minimum number of different products. It is evident that this type of manufacturing process will be less costly and will lead to the production of a more uniform end product. On this basis, more information of a design nature should be available pertaining to the standard sizes, and this will ultimately be of value to the core user.

The objectives to be met by the standard, therefore, are to provide a sufficient range of core dimensions so that the majority of typical circuit design problems can be handled directly by one or another listed size. The list must also be small enough to gain the advantages mentioned of quality and economy.

Paper 57-206, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 23, 1956; made available for printing December 11, 1956.

The personnel of the Core Size Standardization Working Group of the Materials Subcommittee of the AIEE Magnetic Amplifiers Committee are: R. W. Roberts, *chairman*, H. R. Bosworth, D. O. Schwennesen, L. W. Buechler, C. C. Horstman, H. L. Kellogg, C. W. Lufcy, J. R. Mitch, F. F. Siebert, and C. B. Wakeman.

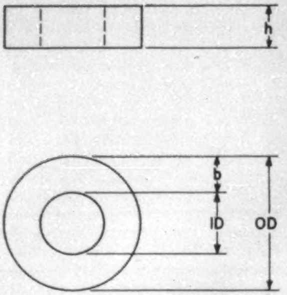


Fig. 1. Toroidal core dimensions

Procedure

Initially, the efforts of the group were applied to the investigation of theoretically optimum core configurations for typical circuit applications. It was felt that if a list of standard sizes was recommended, which was later found to differ widely from optimum, the standard would be of little lasting value. Although this initial work led to definite results,[2] it was found that few general conclusions could be reached as to particular optimum configurations. Furthermore, there was no evidence to indicate that the majority of cores which were in wide usage throughout the industry would soon become obsolete because of wide variances from any theoretical optimum.

Based on these considerations, the group turned to the task of establishing a restricted listing of core sizes following some logical progression system, and containing as far as possible those sizes currently in wide usage throughout the industry. The progression system chosen was that of approximate power-handling capacity. It can be shown that, for designs in which the power or output winding occupies the majority of the available winding area, that power output is proportional to the product of window and core cross-sectional areas A_cA_w. The degree to which this series was followed in the final selection can be seen in Table I.

Although the initial efforts of the group led to no indications of a particular optimum-core configuration, information was obtained that the *ID/OD* ratio (ratio of inside to outside diameters) has a major influence on core performance. In some applications it appears that a high value of *ID/OD* ratio is desirable in order to achieve maximum amplification and minimum response time, while in other applications a lower value of *ID/OD* ratio appears desirable to achieve maximum power output for minimum size and weight. Consequently, it was decided to establish two separate series of sizes, one with higher and one with lower values of *ID/OD* ratio.

The next factors considered are the values of strip width and inside diameter. A considerable degree of uniformity is already in existence with respect to these factors, and any major change would be prohibitively costly because of the tooling changes required. This situation was recognized by setting down listings of standard values for strip width and inside diameter, as given in Table II.

The final factor considered was the *h/b* (ratio height to buildup ratio). On the basis of experience, it appeared that *h/b* ratios of less than one were very seldom, if ever, required from a performance standpoint, while *h/b* ratios of greater than approximately four were approaching the impractical, from a core-case design standpoint. Thus, a limitation of the *h/b* ratio to values between one and four was chosen as being practical for the greater majority of applications.

Once these governing factors were established and agreed upon, the procedure followed was to calculate the important parameters for a complete matrix of core sizes, starting with the standard values of strip width and inside diameter. More than 100 different core sizes appeared in this matrix. Finally, using the matrix as a base, and a listing of core sizes in usage throughout the industry, particular sizes were chosen which would best fit into a smooth progression. The results are given in Table I.

The recommended core sizes set forth in this report will, in the opinion of the working group, be satisfactory for the majority of applications in magnetic amplifiers and associated devices. It is hoped that the industry as a whole will benefit through resulting improvements in quality and reductions in operating costs.

Table I. Proposed Standard Dimensions

ID×OD×h	A_w, In^2	A_c, In^2	A_cA_w, In^4	l, Cm	h/b	ID/OD
Series 1, Low Gain						
1/2×3/4×1/8	0.196	0.0157	0.00307	4.92	1.00	0.667
5/8×1×3/16	0.307	0.0352	0.0108	6.36	1.00	0.625
$3/4\times1\frac{1}{8}\times3/16$	0.441	0.0352	0.0155	7.38	1.00	0.667
$1\times1\frac{1}{2}\times1/4$	0.785	0.0625	0.0491	9.84	1.00	0.667
$1\frac{1}{4}\times2\times3/8$	1.22	0.141	0.172	12.7	1.00	0.625
$1\frac{1}{2}\times2\frac{1}{2}\times1/2$	1.77	0.250	0.442	15.8	1.00	0.600
2×3×1/2	3.14	0.250	0.785	19.7	1.00	0.667
2×3×1	3.14	0.500	1.57	19.7	2.00	0.667
$2\frac{1}{2}\times3\frac{3}{4}\times1\frac{1}{2}$	4.90	0.936	4.59	24.9	2.40	0.667
$3\frac{1}{4}\times5\times1\frac{1}{2}$	8.30	1.31	10.9	32.4	1.71	0.650
4×6×2	12.6	2.00	25.2	39.4	2.00	0.667
Series 2, High Gain						
1/2×5/8×1/8	0.196	0.00783	0.00154	4.47	2.00	0.800
0.650×0.900×1/8	0.332	0.0156	0.00519	6.20	1.00	0.722
3/4×1×1/4	0.441	0.0312	0.0138	6.93	2.00	0.750
$1\times1\frac{3}{8}\times1/4$	0.785	0.0469	0.0368	9.39	1.33	0.726
$1\frac{1}{4}\times1\frac{3}{4}\times1/2$	1.22	0.125	0.1529	11.9	2.00	0.714
$2\times2\frac{1}{2}\times1/2$	3.14	0.125	0.392	17.9	2.00	0.800
$2\frac{1}{2}\times3\frac{1}{2}\times1/2$	4.90	0.250	1.23	23.7	1.00	0.714
$2\frac{1}{2}\times3\frac{1}{2}\times1$	4.90	0.500	2.45	23.7	2.00	0.714
$3\frac{1}{4}\times4\frac{1}{2}\times1\frac{1}{2}$	8.30	0.936	7.77	30.6	2.40	0.722
$4\times5\frac{1}{4}\times2$	12.6	1.25	15.7	36.7	3.20	0.761

Maximum allowable strip gauge to be used for any of the listed core sizes is given by the relationship, strip thickness $\leq ID/125$.

Core-Size Standards

SCOPE

The scope of this standard is restricted to the recommendation of nominal dimensions for toroidal strip-wound cores made from magnetic tape, between approximately 0.014 and 0.0005 inch in thickness, and without regard to the type of magnetic material used. Bobbin-wound cores of ultrathin magnetic tapes are not included. Also excluded from consideration are recommendations on dimensional tolerances, space factors, number of wraps of tape, core-case shape or size, and particular values of strip thickness.

Table II. Proposed Standards for Choice of Future Core Dimension

Strip Widths, In	Inner Diameters, In
1/8..3/16..1/4	1/2..5/8..3/4
3/8..1/2	$1....1\frac{1}{4}..1\frac{1}{2}$
$3/4..1....1\frac{1}{4}$	$2....2\frac{1}{2}........3\frac{1}{4}$
$1\frac{1}{2}...2$	4....5....6
4	8

ID/OD ratio: $0.5 \leq ID/OD \leq 0.85$; height/buildup ratio: $1 \leq h/b \leq 4$; strip thickness: $\leq ID/125$.

Table III. Commonly Used Core Dimensions Not Included as Proposed Standards

ID×OD×h	A_w, In²	A_c, In²	A_cA_w, In⁴	l, Cm	h/b	ID/OD
$5/8\times1\times1/4$	0.307	0.0468	0.0144	6.36	1.33	0.625
$1\times1^1/_4\times1/8$	0.785	0.0156	0.01225	8.94	1.00	0.800
$1\times1^1/_4\times1/4$	0.785	0.0312	0.0245	8.94	2.00	0.800
$1\times1^1/_2\times3/8$	0.785	0.0937	0.0735	9.84	1.50	0.667
$1\times1^1/_2\times1/2$	0.785	0.125	0.0981	9.84	2.00	0.667
$1^1/_4\times1^3/_4\times1/4$	1.22	0.0625	0.0763	11.85	2.00	0.715
$1^5/_8\times2\times1/4$	2.07	0.0468	0.097	14.47	1.33	0.813
$2^1/_2\times3\times1/2$	4.90	0.125	0.613	21.73	2.00	0.833
$2^1/_2\times3^3/_4\times1^1/_4$	4.90	0.7812	3.83	24.95	2.00	0.667

DEFINITIONS OF TERMS AND SYMBOLS

The symbols used in definition of core dimensions are shown in Fig. 1.

ID = inner diameter of the core
OD = outer diameter of the core
h = width of the strip from which the core is wrapped, which corresponds to the core height
b = core build-up, $b = OD - ID/2$
A_c = core area without regard to tape-stacking factor; $A_c = b \times h$
A_w = window area $A_w = \pi ID^2/4$
l = magnetic path length calculated as the arithmetic mean; $l = \pi(ID+OD)/2$

STANDARD CORE SIZES

In Table I are shown the core sizes recommended as standard. This listing is divided into two groupings based on values of ID/OD ratio. Cores in the first grouping, entitled "Low Gain," have ID/OD ratios of less than 0.70. These cores are generally best suited for applications where considerations of maximum output power for minimum size and weight are of prime importance. Cores in the second grouping, entitled "High Gain," have ID/OD ratios greater than 0.70. These cores are generally best suited for applications where maximum power gain and/or minimum response time are of prime importance.

The maximum tape thicknesses which may be used for the various recommended standard core sizes are subject to a natural limitation because of practical winding difficulties. A recommended standard maximum tape thickness, noted in the footnote of Table I, is 1/125 of the ID for the core. This recommendation applies to both standard and non-standard cores sizes.

NONSTANDARD CORE SIZES

It is anticipated that in a few very critical applications it will be impossible to utilize one of the recommended standard core sizes. For this reason a tabulation of core sizes which are not recommended as standard, but which have found wide usage is given in Table III. This tabulation is intended as a first guide in the choice of a non-standard core size.

If in the solution of a critical design problem it is impossible to use a core size listed in Table I or in Table III, it is strongly recommended that general standards on strip width and inner diameter be adhered to. These standards, along with recommendations as to limiting values of h/b ratio, ID/OD ratio, and strip thickness are given in Table II.

References

1. STANDARDS FOR THE PRESENTATION OF DATA ON MAGNETIC AMPLIFIER CORE MATERIALS, AIEE Committee Report. *AIEE Transactions*, vol. 74, pt. I, Nov. 1955, pp. 598–99.

2. INFLUENCE OF ID/OD RATIO ON STATIC AND DYNAMIC MAGNETIC PROPERTIES OF TOROIDAL CORES, R. W. Roberts, R. I. VanNice. *Ibid.*, pp. 599–607.

Circuit for a Magnetic Subharmonic Pulser

K. C. HU
ASSOCIATE MEMBER AIEE

Y. H. KU
FELLOW AIEE

THIS PAPER describes a saturable reactor circuit from which pulses of the subharmonic order of the frequency of the applied voltage are produced. The basic idea consists of resetting the flux in the core of a half-wave self-saturating magnetic amplifier by means of a discharge current from a capacitor which has been previously charged during the saturation period. The effect of the discharge current can be adjusted by varying a series resistor so that it will last for a desired number of cycles before the reactor can be saturated again. In this paper, the mechanism of operation of such circuitry is explained and experimental results are given using 60-cycle frequency. The results are partly checked by analysis. This device, because of its simplicity, could be useful as a subharmonic generator or a frequency divider provided that the input frequency is known.

Circuit Description

The basic circuit is shown in Fig. 1. The saturable reactor has two windings, the gate winding N_g and the reset winding N_r. The resistor R is common to both the reset and gate circuits. The value of the capacitive reactance of C, X_c, is about the same order as the load resistance R_L but is much smaller than R.

It is seen from Fig. 1, before saturation, that the entire input voltage would appear across gate winding N_g as the inductance of the coil is large. Upon saturation, if the winding resistances are negligible, R is very large compared with X_c, and the post-saturation inductance of the coil is negligible, then a conduction current will flow through the capacitor, the rectifier, and the load resistance, and its magnitude will be limited only by their combined impedance. Actually, as R is not infinite, it shunts part of the conduction current away from C. When C discharges in the reset circuit, R is

---

Paper 57-99, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted September 25, 1956; made available for printing November 21, 1956.

K. C. HU is with the Brown Instrument Division, Minneapolis-Honeywell Regulator Company, Philadelphia, Pa., and Y. H. KU is with the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa.

Acknowledgment is due to I-T-E Circuit Breaker Company for providing the Permeron tape-wound core used in the experiments. This paper is based on a thesis submitted by K. C. Hu to the Moore School of Electrical Engineering, University of Pennsylvania, in partial fulfillment of the requirement for the degree of Master of Science in electrical engineering.

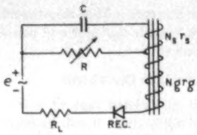


Fig. 1. Circuit diagram

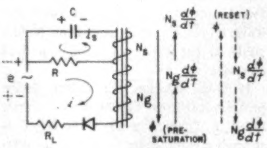


Fig. 2. Relative polarity of induced voltages

also a factor in determining the rate of discharge. Its exact function will soon be seen.

Physics of Operation

The physics of operation can be explained best by observing the change of capacitor voltage in successive half-cycles after a steady state is reached. The change of capacitor voltage depends greatly on the mutual inductance between N_g and N_s. There is an induced voltage in one winding when magnetizing current flows in the other. The relative polarity of the induced voltages in N_g and N_s is shown in Fig. 2. The induced voltage in N_s by current i in N_g during positive half-cycles is in the direction of discharge of the capacitor. Hence it is to increase the rate of discharge of C. The voltage induced in N_g by discharge current in N_s in the negative half-cycles is in the direction of conduction of the rectifier. It is opposed only by the negative-half-cycle applied voltage.

After a steady state is reached, the core will saturate in a positive half-cycle designated as *0–1* in Fig. 3. At the instant *O* or the beginning of the half-cycle, the capacitor has a certain voltage V_0, and the flux in the core is at a certain level ϕ_0 which is determined through previous intervals. At the instant *f* where the core saturates, the flux has changed from ϕ_0 to the saturation value ϕ_s. In the mean time, a voltage is induced in N_s between the instants *O* and *f* to discharge the capacitor to a still lower voltage. Upon saturation or firing, the entire voltage at the instant *f* would appear across R_L in series with C because the reactor inductances drop to practically zero, the winding resistances and the rectifier forward resistance are negligible, and the value of R is relatively large. The instantaneous conduction current is nearly equal to the applied voltage at instant *f* divided by the resistance *RL*.

Next, a voltage will soon develop across C to oppose the applied voltage which is decreasing in the latter part of the positive half-cycle. The conduction will then stop at an instant *a* as shown in Fig. 3 at which the capacitor voltage due to current-time integral between *f* and *a* equals the applied voltage at *a*.

During the remaining part of the first half-cycle, *a-1*, the capacitor will start to discharge in the reset circuit. Initially, the discharge current is limited only by R as the reactor is still in the saturated condition. When the flux starts to reset due to the negatively applied ampere-turns, however, the discharge current is limited by the magnetizing current requirement associated with the flux change. In other words, it is limited by the width of the hysteresis loop along which the core is operating under given conditions. The core nevertheless can hardly reset at this interval because there is no reversed voltage available in the gate circuit to oppose the induced voltage in N_g.

The flux starts to reset during *1–2*, the first negative half-cycle. The reversed applied voltage prevents any circulating current from flowing in the gate circuit and at the same time the rectifier prevents any current from flowing in the opposite direction. At the end of this period, instant *2*, the flux is reset to ϕ_2.

During the interval *2–3*, the second positive half-cycle, the applied voltage takes over again to change the flux toward ϕ_s. In doing so, a voltage is induced in N_s to accelerate the discharge of C. The increase of the discharge current, however, must be supplied by the voltage source by the relation of equal ampere-turns between N_s and N_g. The flow of such current, especially when the capacitor voltage is considerable, means a large voltage drop across R and that much less voltage will appear across N_s, the volt-time integral of which determines the amount of flux change toward saturation during this half-cycle. It is therefore very possible that even though the applied voltage is higher than the voltage that the reactor is able to absorb, the core cannot be saturated.

For the remaining intervals, the events will repeat themselves but the amount of reset will be less and less and the change toward ϕ_s more and more as V_c decreases gradually. If, at instant *O*, the end of an integral number of cycles, the flux level and the capacitor voltage could be returned to ϕ_s and V_0, respectively, the exact values as at the beginning, then the core would fire again at the exact instant in the next positive half-cycle. Fig. 3 shows the case of a fourth subharmonic cycle,

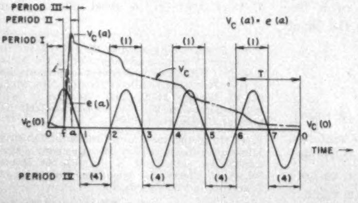


Fig. 3 (above). Variation of capacitor voltage in a fourth subharmonic cycle

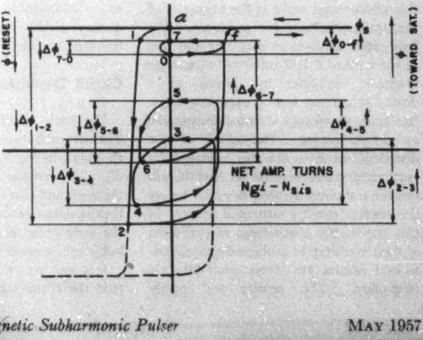


Fig. 4 (right). Cycling of flux in fourth subharmonic cycle

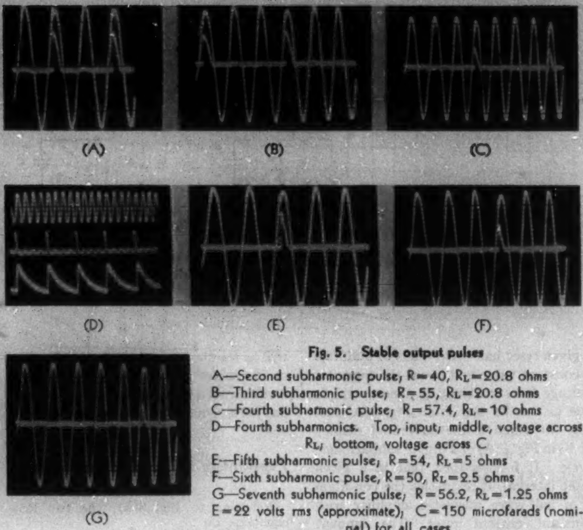


Fig. 5. **Stable output pulses**

A—Second subharmonic pulse; $R=40$, $R_L=20.8$ ohms
B—Third subharmonic pulse; $R=55$, $R_L=20.8$ ohms
C—Fourth subharmonic pulse; $R=57.4$, $R_L=10$ ohms
D—Fourth subharmonics. Top, input; middle, voltage across R_L; bottom, voltage across C
E—Fifth subharmonic pulse; $R=54$, $R_L=5$ ohms
F—Sixth subharmonic pulse, $R=50$, $R_L=2.5$ ohms
G—Seventh subharmonic pulse; $R=56.2$, $R_L=1.25$ ohms
$E=22$ volts rms (approximate); $C=150$ microfarads (nominal) for all cases

where the reactor is saturated once in four complete cycles of the frequency of the applied voltage. Fig. 4 depicts the cycling of flux during the subharmonic cycle. The numbers in Fig. 4 correspond to the instants marked on the time axis of Fig. 3.

Experimental Results

Based on the foregoing description, the requirement of the reactor is that the magnetizing current should be very small, and that it should have extremely high inductance before saturation and practically zero inductance as soon as saturation is reached. Square *B-H* loop material hence would be most suitable for this purpose. Permeron* tape-wound core was used in the experiments. The core, $1\frac{1}{2}$ inches inner diameter and 2 inches outer diameter, 20 millimeters wide, and 0.002 inch tape thickness, is wound with 438 turns of N_g and 120 turns of N_s.

With a General Electric germanium rectifier, *GE4JA211AH1AC2*, subharmonic pulses across R_L of the order up to the tenth are obtainable from this reactor by varying other circuit parameters. Stable output pulses of the order from the second to the seventh are recorded by cathode-ray oscilloscope as shown in Figs. 5(A) to 5(F). The values of the circuit constants and the applied voltage are indicated in each case. It is to be noted that there are numerous combinations of R, R_L, and the applied voltage that would give a specific order of subharmonic pulses. The value of C is not critical.

Following the foregoing description, the decreasing rate of capacitor voltage in a positive half-cycle is always higher than that in a negative half-cycle because the discharge is accelerated by an induced voltage in N_s. This is verified in Fig. 6(A) which shows V_c as compared with the applied voltage in an actual fourth subharmonic cycle. It can be seen from Fig. 6(A) that V_c decreases almost along a straight line in the negative half-cycles, i.e., the capacitor discharges at a constant rate in the reset period (also refer to Fig. 9(A) and Fig. 10).

Figs. 6(B) and 6(C) show the actual flux cycling in the core in a fourth and seventh subharmonic cycle. These displays are obtained by an arrangement as shown in Fig. 7. The vertical input to the oscilloscope is obtained through a third winding and an integrating network in a usual manner. The horizontal input is provided by the difference of voltage drop across the resistors R_1 and R_2. The ratio of R_1/R_2 is equal to N_g/N_s. The input $R_1i-R_2i_s$ is thus proportional to the net ampere-turns on the core $N_gi-N_si_s$. The gradual shifting of the loops to the right in Fig. 6(B) may be due to the fact that R_1/R_2 was not chosen to equal to N_g/N_s exactly.

Conclusions and Discussions

If the *B-H* loop of the core material could be idealized as perfectly rectangular, then the afore-mentioned events could be separated into four successive intervals. Each interval could be represented by a simple equivalent circuit with further assumptions of perfect rectifier and negligible winding resistances. The four intervals are as follows

1. *Period I: Presaturation.* This includes the interval *O-f*, as referred to Fig. 3, in the firing half-cycle and all the other positive half-cycles.
2. *Period II: Conduction.* The interval when the core is saturated corresponding to *f-a* in Fig. 3.
3. *Period III: Starting of Reset.* Period *a-1* in Fig. 3 when the reset current starts to flow while the gate voltage is still positive.
4. *Period IV: Reset.* It includes all the negative half-cycles.

The equivalent circuits for these periods are given in Figs. 8(A) to 8(D).

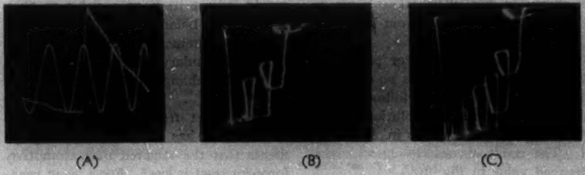


Fig. 6. **Flux cycling**

A—V_c in a fourth subharmonic cycle
B—Flux cycling in a fourth subharmonic cycle
C—Flux cycling in a seventh subharmonic cycle

---

* I-T-E Circuit Breaker Company trade-mark.

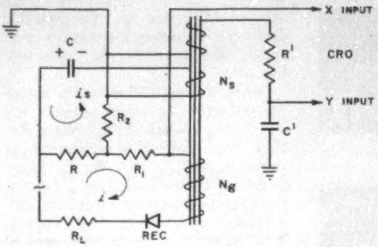


Fig. 7. Circuit for obtaining display of Fig. 6(B)

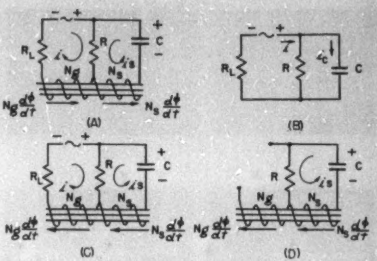


Fig. 8. Equivalent circuits

A—Period I B—Period II C—Period III D—Period IV

In the reset period, further assumption is made that the induced voltage in N_g is less than the reversed applied voltage at all times.

When the flux is changing toward saturation under the positive applied voltage the amount of flux change can always be related to the volt-time integral which appears across the winding terminals. For square *B-H* loop material like Permeron, the magnetizing current required for the flux change is small. In period I, as stated before, the applied voltage has to supply the capacitor discharge current in addition to the magnetizing current. In the presence of the discharge current, especially when the capacitor voltage is considerable, the magnetizing current could be neglected for practical purposes. Under these conditions, the circuit of Fig. 8(A) can be analyzed mathematically without the requirement of detailed information on the dynamic characteristics of the core material.[1] Analytic solution for the capacitor voltage V_c and the discharge current i_s is given in Appendix I. Fig. 9(A) shows a calibrated display of V_c, i_s, and i_c, the conduction current in a second subharmonic cycle. The variation between instants *2* and *3* is checked by numerical solution from Appendix I as shown in Fig. 9(B). The value of V_c at instant *2* is taken from Fig. 9(A) for calculation.

The conduction period, or period II, shown in Fig. 8(B) presents no difficulty for analytic formulation following the foregoing assumptions. Solution of the conduction angle is given in Appendix II. The result checks closely with the observed value in Fig. 9(A).

The analytic presentation of Figs. 8(C) and 8(D) is more difficult and will not be given here. From observation, it seems that the discharge current in a given reset half-cycle tends to maintain a constant value and hence the capacitor voltage decreases along an almost straight line as can be seen from Fig. 9(A) and Fig. 10 which is the calibrated display of V_c in Fig. 6(A) of a fourth subharmonic cycle. It follows then that in a given reset interval, it resembles the condition of a half-wave magnetic amplifier with constant current reset. The amount of reset in that period as well as the rate of reset would then depend upon the dynamic behavior of the core material. This implies that the amount of reset would depend on the core geometry, the tape thickness, material heat treatment, and its previous magnetization history.[2] The fact that the core is not always reset from saturation level tends to make any mathematical analysis even more difficult.

It could be said qualitatively with reference to the experimental results that for a core reset under the influence of capacitor discharge, the reset would take place along a specific dynamic *B-H* loop which is determined by the capacitor voltage at the beginning of the reset period and the series resistance *R*. A higher V_c at the beginning or smaller *R* would result in a wider loop. The flux would reset at a faster rate under such condition and the discharge current would adjust itself to a value so that it would be balanced out by the eddy-current requirement associated with the domains boundary velocities as required by the flux change along that particular loop.[3]

For a reset from partially saturated condition, it has been shown in experiments that the required magnetomotive force for reset is less.[2] For a given reset magnetomotive force, the flux would reset at a higher rate than that from a completely saturated condition. This phenomenon perhaps is the very basic reason for the operation of this circuitry as it enables the capacitor to reset the core flux even though the residual voltage has been reduced to a very low value.

Appendix I. Analysis of Period I

Referring to Fig. 8(A), the two loop equations are

$$e=E\sin\omega t=R(i+i_s)+N_g d\phi/dt+iR_L \quad (1)$$

$$N_s d\phi/dt=\frac{1}{C}\int i_s dt+R(i+i_s)+ri_s-V_{c(o)} \quad (2)$$

where r is the resistance inserted in reset circuit for measuring i_s.

The following differential equation is obtained by combining the foregoing equations and using the relation $N_g i=N_s i_s$.

$$N^2/C\int i_s dt+ai_s=NE\sin\omega t+N^2V_{c(o)} \quad (3)$$

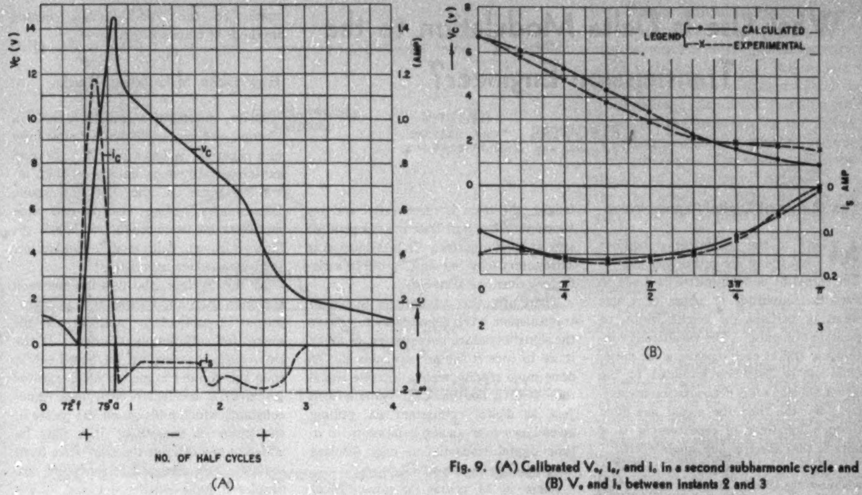
where $N=N_g/N_s$, $a=R(1+N)^2+R_L+N^2r$, and $V_{c(o)}$ is the capacitor voltage at the beginning of the period. The solution of equation 3 is

$$i_s=\frac{N^2}{a}\left(V_{c(o)}-\frac{N\omega EaC}{N^4+\omega^2a^2C^2}\right)e^{-\frac{N^2}{aC}t}+\frac{\omega EC}{V_{c(o)}(N^4+\omega^2a^2C^2)^{1/2}}\sin(\omega t+b) \quad (4)$$

where $b=\pi/2-\tan^{-1}\omega aC/N^2$. The capacitor voltage at any instant v_c is

$$v_c=V_{c(o)}-1/C\int_0^t i_s dt =\left(V_{c(o)}-\frac{aN\omega EC}{N^4+\omega^2a^2C^2}\right)e^{-\frac{N^2}{aC}t}+\frac{aN\omega EC\cos\omega t-N^2E\sin\omega t}{N^4+\omega^2a^2C^2} \quad (5)$$

The calculated curve in Fig. 9(B) is obtained by substituting the following constants into equations 4 and 5. $N=3.57$, $w=377$ radians per second, $E=27.2$ volts peak, $C=187.5$ microfarads, $R=36.8$ ohms, $R_L=12.65$ ohms including resistance in N_g, $r=2.96$ ohms including resistance in N_s, $V_{c(o)}=6.8$ volts by observation from Fig. 9(A).



(A)

(B)

Fig. 9. (A) Calibrated V_c, I_c, and I_s in a second subharmonic cycle and (B) V_c and I_s between instants 2 and 3

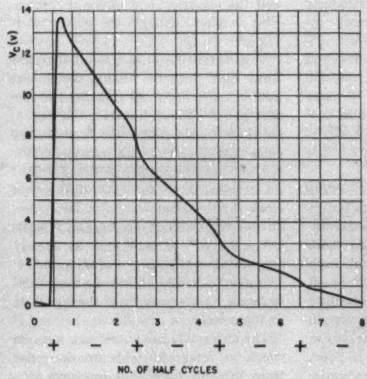


Fig. 10. Calibrated display of V_c in a fourth subharmonic cycle

Appendix II. Analysis of Period II

Referring to Fig. 8(B), the loop equations are

$$e = E\sin(\omega t+f) = R_L i + \frac{1}{C}\int i_c dt \tag{6}$$

$$R(i-i_s) = \frac{1}{C}\int i_c dt \quad (r \text{ neglected}) \tag{7}$$

where f is the firing angle and the capacitor voltage is assumed to be zero at the instant of firing. The following differential equation is then obtained.

$$\frac{R_1}{C}q + R_L dq/dt = E\sin(\omega t+f) \tag{8}$$

where q is instantaneous charge on the capacitor. Solving,

$$q = CE\frac{\omega CR_L\cos f - R_1\sin f}{R_1^2+\omega^2C^2R_L^2}e^{-\frac{R_1}{CR_L}t} + \frac{CE\sin(\omega t+k)}{(R_1^2+\omega^2C^2R_L^2)^{1/2}} \tag{9}$$

where $k = f - \tan^{-1}\omega CR/R_1$ and $R_1 = (R_L+R)/R$.

The following relation must hold if conduction stops at instant a. The difference between a and f is the conduction angle c.

$$\mathrm{Sin}(f+c)(R_1^2+\omega^2C^2R_L^2) = (\omega CR_L\cos f - R_1\sin f)e^{-\frac{R_1 c}{\omega CR_L}} + R_1\sin(f+c) - \omega CR_L\cos(f+c) \tag{10}$$

Solving equation 10 by cut-and-try, one obtains $c = 77.5$ degrees while the observed value is 79 degrees as shown in Fig. 9(A).

References

1. Dynamic Core Behavior and Magnetic Amplifier Performance, L. A. Finzi, D. L. Critchlow. *AIEE Transactions*, vol. 76, pt. I, May 1957, pp. 229–40.

2. Flux Resetting Characteristics of Several Magnetic Materials, Hoobert Huhta. *Ibid.*, vol. 73, pt. I, May 1954, pp. 111–14.

3. The Influence of Magnetic Amplifier Circuitry Upon the Operating Hysteresis Loops, Harold W. Lord. *Ibid.*, vol. 72, pt. I, 1953 (Jan. 1954 section), pp. 721–28.

What Use Is Delta Modulation to the Transmission Engineer?

F. K. BOWERS
NONMEMBER AIEE

The Merits of Digital Transmission

MOST of the transmission methods used today are rather elementary. The signal to be transmitted is made to vary the amplitude or phase of a sine wave, or perhaps the height, width, or position of a pulse. The variation is continuous, and at each repeater some noise, cross-talk or distortion is added to the signal, causing some deterioration in quality. By the time the signal has gone through hundreds of repeaters, it is a miracle that there is any quality left.

What is clearly needed is some device whereby the effects of noise and distortion in one link are not passed on to the next one. This can be done if, instead of allowing the amount of modulation to vary continuously, it is restricted to a discrete number of levels. If, for example, the height of any pulse sent out is one of four possible values: 0, 1, 2, or 3 volts, then it should not be hard for the next repeater to decide which of the four values to assign to any incoming pulse, and to generate new pulses of the exact heights required. This is possible as long as the peak of noise and distortion stays below some limit: as long as this happens, the signal can pass through any number of repeaters without change.

The simplest digital system to build is one where there are just two possible amplitude levels for each pulse, such as 0 and 1 volt. A repeater can then consist of just one pulse generator (such as a blocking oscillator) under the control of a master clock. At regular intervals, determined by the clock, the pulse generator is able to send out a standard size pulse, but it will do so only if triggered by an incoming pulse greater than some threshold value. A train of binary pulses is sent from one repeater to the next and, as long as all the distortions and interference do not add up to more than half the pulse height, the train is regenerated and retimed to the extent that it is as good as new at each repeater. If transmission is satisfactory over one link, it will be satisfactory over one hundred.

There are other advantages to binary transmission. If it is necessary to process the signal to reduce redundancy, to delay it, or to code it for privacy, this can be done more readily with a discrete signal than with a continuously variable one. Just as digital computers are gaining ascendancy over analog computers, so in time digital transmission may displace many analog systems.

There is, of course, a price. Most signals start out as continuously variable signals, and there is the trouble of coding these to convert them into binary form. This conversion costs money, introduces some quantizing noise, and greatly increases the bandwidth needed for transmission. The quality of the transmission band may indeed be very low, but the quantity required is enormous: perhaps seven times the original bandwidth.

Delta Modulation and PCM

Delta modulation is one way of coding a continuously variable signal into binary form. Another, somewhat better known, is pulse code modulation, or PCM. Both start with the same input signal, both convert this into a stream of binary, on-off pulses, and both eventually reconstitute a signal which has, it is hoped, a certain likeness to the input. They differ only in the way in which the coding is done. The arrangement of the pulses for a particular input signal will be quite different in the two systems and, unless the right kind of decoder is in use, interpretation will be impossible.

The two systems are therefore rivals for patronage and the decision between them will depend on their quality and their price. (A recent comparison on more theoretical grounds was given by Zetterberg.[1]) In this paper delta modulation is examined and an attempt made to determine what its transmission quality is, how this compares with PCM, and how much equipment is needed for the two systems. No further reference will be made to the old-fashioned analog methods of transmission.

How Delta Modulation Works

Delta modulation was invented in Europe, and has been described most fully in a paper by de Jager in 1952.[2] Several experimental systems have been built, including some at the Bell Telephone Laboratories. These incorporated some modifications to be mentioned later. As far as is known, delta modulation has not yet been used commercially.

In this system, the relation between the pulse train and the signal to be transmitted is particularly simple, and this makes for very simple decoders. If the two values of pulses are $+1/2$ and $-1/2$, then the decoder is merely an integrator (such as a condenser with long time-constant) which adds up all the pulses in the train. A smoothing filter may be added to smooth out the step wave form so obtained. Since the repeaters are simpler if pulse-heights $+1$ and 0 are used, it is customary to add at the receiver a constant negative current which subtracts during every pulse interval half the charge brought by a $+1$ pulse. This, in effect, changes the pulses to $+1/2$ and $-1/2$.

It is therefore easy to convert the pulse train back into the analog signal, but how is the pulse-train generated in the first place? This is illustrated in Fig. 1. On the left is the transmitter or coder, the chief element of which is a pulse generator g, which has two possible modes of action: At regular intervals, controlled by a clock, it either sends out a standard pulse, or does not send out a pulse, i.e., it sends out a $+1$ or a 0 pulse and, as already seen, this is equivalent to sending out $+1/2$ or $-1/2$. The choice of action is determined by the polarity of the signal at its input terminal.

The coder also has a feedback loop, in which an integrator adds up the pulse train $p(t)$, to form the step-wave $v(t)$. This integrator is a replica of the one found at the receiver, and the received wave form $v(t)$ is therefore available at the transmitter. Then, $v(t)$ is compared with the incoming signal $u(t)$ and the difference, $e(t)$, is fed to the pulse generator to determine the type of pulse next sent out.

Therefore, if at any time the input signal u is larger than the reproduced signal v, then the next pulse sent will be positive and will increase v. If, on the other hand, u is smaller than v, the next pulse is negative and v decreases. In this way, the output follows the input quite

---

Paper 57-136, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted November 19, 1956; made available for printing November 23, 1956.

F. K. Bowers is with Bell Telephone Laboratories, Inc., Murray Hill, N. J.

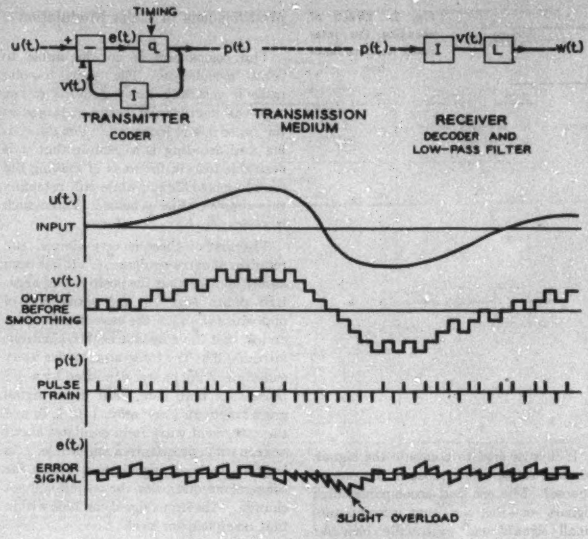


Fig. 1. Simple delta modulation

closely. Except when the system is overloaded, the error voltage does not get much larger than one quantum or one unit step (the effect of one positive pulse). In the absence of an input signal, the pulse train is an endless sequence of plus and minus pulses, and the reproduced signal a square wave, one quantum peak-to-peak, and of half the pulse frequency.

PCM Coding and Decoding

By comparison with delta modulation, PCM is a much more complicated and more sophisticated transmission scheme.[3] It has had considerable experimental application, and its plan of operation will merely be summarized here.

If the incoming signal contains components up to frequency f_0, then it is first sampled at a rate at least $2f_0$. This yields a set of pulses continuously modulated in amplitude. Then, the amplitudes of these pulses are quantized, and assigned to one of a finite number of levels. Typically there might be 128 such levels. Finally, the level number is coded into binary notation, e.g., $9 = 0001001$, and a corresponding set of seven pulses is sent out.

At the receiver, the group of pulses belonging to one sample are sorted out, given binary weighting, and added to reconstitute the sample. The samples are then smoothed with a low-pass filter to give a reconstructed signal.

The quantizing and coding into binary from are often done in a special kind of cathode ray tube,[4] but they can also be carried out with conventional elements, as will be seen later.

Comparing Transmission Quality

Both the two systems described have some maximum signal power that they can handle without overloading, and both have some imperfection in transmission, which is called quantizing noise, which is of a peculiar type: It is not thermal noise, or interference accumulated in transmission. Quantizing noise is produced at the transmitter by the process of forcing a continuous signal to move in finite size steps. The smaller the steps, the less will be the quantizing noise. On the other hand, the smaller the steps, the greater will be the pulse rate required for transmission. Since this pulse rate is usually strictly limited, a reasonable amount of quantizing noise will have to be tolerated.

Quantizing noise is defined as the difference between the input and the output wave form, and therefore includes what is normally considered distortion. However, it often sounds like noise, and as soon as some kind of maximum signal power and some kind of noise power are defined, the temptation is irresistible to combine them into a signal-to-noise ratio, and to use it as an index of transmission quality.

Table I shows such signal-to-noise ratios, but the tabulation is shown with misgivings, and should be viewed with great distrust. Some reasons for this distrust will now be given. For PCM, there is no great difficulty. The maximum signal power is fixed by the capacity of the coder, and the amount of noise and its spectrum were computed in 1948 by Bennett.[5] For most purposes, the noise sounds much like thermal noise.

Delta modulation, on the other hand, has no fixed maximum amplitude. It overloads when the slope of the signal gets too large and the biggest slope the system can reproduce is one chang-

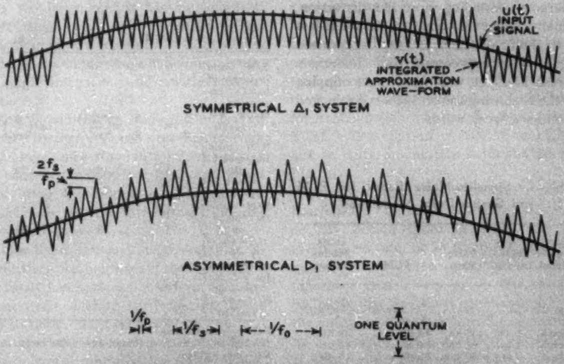


Fig. 2. Wave forms of symmetrical and asymmetrical delta modulation

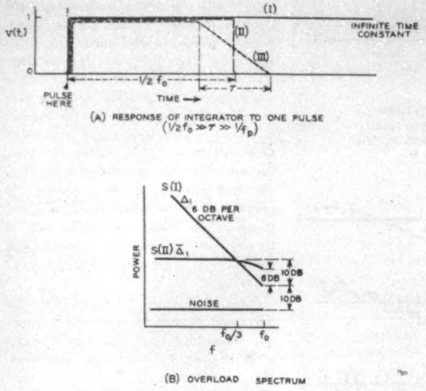


Fig. 3. Effects of changing the integrator response function

ing by one quantum every pulse interval. Consequently, the maximum signal power depends on the type of signal: for a sine wave of frequency f it varies as $1/f^2$. To compute it for speech and TV (television) some rather arbitrary assumptions must be made about these signals. For these comparisons, assume the peak power spectrum of TV is flat, and that of speech has a maximum at 800 cycles per second and falls off 6 db (decibels) per octave up to 4 kc.

The quantizing noise of delta modulation is also peculiar in many respects. When the signal power is large, it is fairly constant and fairly uniform: this is the noise measured by de Jager and others. This noise amplitude may be considerably smaller than one quantum. However, as the signal power is reduced, the noise increases in amplitude and changes in character, sounding more like distortion. Finally, when the signal amplitude is less than half a quantum, the reproduced signal disappears completely. There is a threshold effect, which can cause clipping of the beginning and end of a syllable, and the loss of weak words.

If it is desired to compute the signal-to-noise ratio, which noise is the one to choose? The one that accompanies high signals, or the one that accompanies small signals and eventually obscures them? The latter is probably more interesting: the signal-to-noise ratio then indicates the range of signal powers that may be transmitted, from the highest that just overloads the system, to the lowest that is just equal to noise. This is how the ratios in Table I were derived. They give some means of comparing delta modulation with other systems. Just how accurate the comparison is will have to be checked by subjective listening tests.

Modifications of Delta Modulation

The comparison is not favorable to delta modulation. The signal-to-noise ratio is generally many decibels worse, and the discrepancy becomes larger as the pulse rate is increased. Yet the coding and decoding is so simple that it is desirable to look for ways of making the system more efficient while still retaining this simple coding scheme. Three such modifications have been found.

The first of these is very simple, and requires no extra equipment. It has been assumed so far that the positive and negative pulses sent out are equal and of opposite sign or, in the case of pulses +1 and 0, that the constant negative current introduced at the integrator carries away exactly half the charge deposited by a +1 pulse. In that case, zero input signal gives rise to an alternation 1, 0, 1, 0, and the integrated wave form oscillates about at zero with one quantum amplitude. As long as the input signal stays within the range of one quantum, the output will not change. The input signal can hide within that one quantum level.

It is, however, no more trouble to adjust the coder (or the supply of negative current) so that positive and negative pulses are not exactly equal. In that case, an alternation of pulses might give rise to a steadily rising reproduced signal. However, if the input stays constant at zero, there will be an extra space every now and then, at some frequency f_s, which will bring the output signal down again. After smoothing, the output will now show some components of frequency f_s. If the

Table II. Signal-to-Noise Ratios of Various Delta Systems

			Delta Modulation					
			Single-Stage			Two-Stage		
	PCM	Δ PCM	Δ_1	$\triangleright_1$	$\overline{\triangleright}_1$	Δ_2	$\bar{\Delta}_2$	$\bar{\bar{\Delta}}_2$
S/N ratio in terms of band expansion factor $n = f_p/2f_s$	$\frac{3}{2}2^{2n}$	$\frac{3}{2}2^{2n}\left(\frac{f_s}{\pi f}\right)^2$	$6n^3\left(\frac{f_s}{\pi f}\right)^2$	$\frac{3}{2}n^4\left(\frac{f_s}{\pi f}\right)^2$	$\frac{3}{2}n^4$*	$\frac{3}{2}n^5\left(\frac{f_s}{\pi f}\right)^4$	$\frac{3}{2}n^5\left(\frac{f_s}{\pi f}\right)^2$*	$\frac{3}{2}n^5$†
S/N ratio for speech, db $f_s = 4$ kc $f_p = 56$ kc	44	48	29	40	—	38	44	—
S/N ratio for speech, db $f_s = 4$ kc $f_p = 96$ kc	74	78	33	49	—	50	56	—
f_p required for TV, mc $f_s = 5$ mc S/N = 44 decibels	70	—	—	180	140	—	130	100

S/N = signal-to-noise.
f_s = highest signal frequency; f_p = pulse repetition frequency.
mc = megacycles.
* 4 db less at top of band.
† 8 db less at top of band.

Table I. Signal-to-Noise Ratios for PCM and Simple Delta Modulation

Type of Signal	Bandwidth	Pulse Rate	PCM	Delta Modulation
Sine wave	f_s	f_p	$\frac{3}{2}\times 2(f_p/f_s)$	$\frac{3}{2\pi^2}(f_p/f)^3$
Speech	4 kc.	56 kc.	44 db	29 db
Speech	4 kc.	96 kc.	74 db	33 db
TV	5 mc.	80 mc	52 db	16 db

= frequency of sine wave.

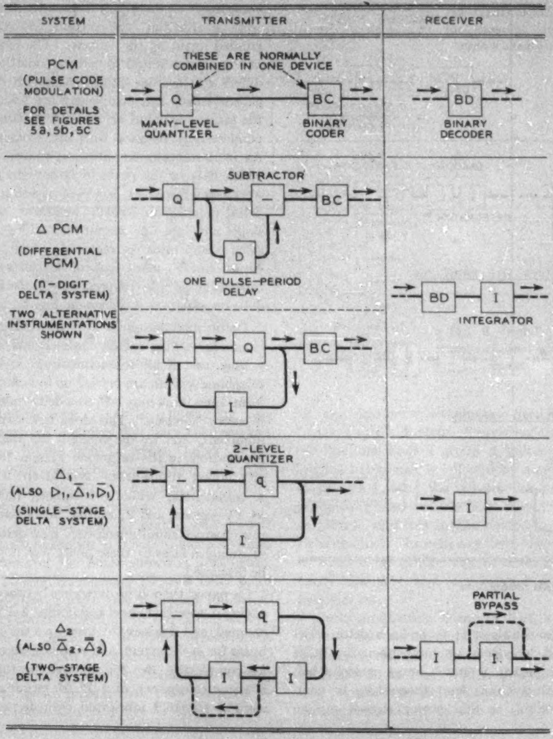

SYSTEM
TRANSMITTER
RECEIVER
PCM
(PULSE CODE MODULATION)
FOR DETAILS SEE FIGURES 5a, 5b, 5c
THESE ARE NORMALLY COMBINED IN ONE DEVICE
Q
BC
MANY-LEVEL QUANTIZER
BINARY CODER
BD
BINARY DECODER
Δ PCM
(DIFFERENTIAL PCM)
(n-DIGIT DELTA SYSTEM)
TWO ALTERNATIVE INSTRUMENTATIONS SHOWN
SUBTRACTOR
Q
−
BC
D
ONE PULSE-PERIOD DELAY
BD
I
INTEGRATOR
−
Q
BC
I
Δ₁
(ALSO ▷₁, Δ̄₁, ▷̄₁)
(SINGLE-STAGE DELTA SYSTEM)
2-LEVEL QUANTIZER
−
q
I
I
Δ₂
(ALSO Δ̄₂, Δ̄₂)
(TWO-STAGE DELTA SYSTEM)
−
q
I
I
PARTIAL BYPASS
I
I

Fig. 4. Showing Δ PCM in relation to PCM delta modulation

discrepancy between positive and negative pulses is large enough, f_s can be made higher than the top frequency in the signal band f_0 and can be eliminated by a final filter.

Fig. 2 shows the behavior of the old symmetrical system and of the new asymmetrical one for a small, slowly varying input. In the new system, the reproduced wave, after smoothing, follows the input much more closely, and the quantizing noise for small signals is reduced. This has been verified experimentally.

Another approach was tried by de Jager. Instead of putting one stage of integration at the receiver, and also in the feedback loop of the coder, there can be two stages in series at each place. Such a 2-stage delta system overloads when the second derivative of the signal gets too large, and it has an improved quantizing noise and threshold effect. It suffers from a tendency to hunt, which is cured by employing a judicious mixture of single integation and double integration.

One would like to know if this can be carried further. Could a small admixture of 3-stage and 4-stage integration perhaps be used in the feedback loop, and would this bring the efficiency up to that of PCM? It is an interesting but unanswered question.

The third and last improvement concerns the exact behavior of the integrators. So far it has been assumed that they are ideal units, adding up incoming pulses for all time. This results in an overload characteristic decreasing with frequency which is quite suitable for some signals. If, on the other hand, the signal has a flat power spectrum, it would be better to make the integrators add over a finite time interval only.

The response of such an integrator to an incoming pulse should be fairly flat for a finite time, about 1/2 f_0, and then decrease to zero in a shorter time. Such a modification of the integrator may also be applied to one or both stages of a 2-stage system.

Fig. 3 shows the various response functions of the integrator. Curve (*I*) is the normal response of an integrator with infinite time-constant, and gives rise to the overload characteristics infinite time-constant, and gives rise to the overload characteristics *S*(*I*); see Fig. 3 (B). If the response is changed to (*II*), the overload point for low frequencies is reduced, but that near the top end is increased, as shown in *S*(*II*). On the other hand, the sharp cutoff of response (*II*) increases the quantizing noise. Response curve (*III*) has about the same overload characteristics as (*II*), but without appreciable increase in noise.

Another Comparison of Transmission Quality

After this extensive doctoring, another comparison with PCM may be faced. Table II shows the signal-to-noise ratios for various types of signals: for PCM, for single-stage delta systems (ordinary and asymmetrical), and for 2-stage systems. Limited integrator response is applied where it is useful, and is indicated by a bar over the Δ.

Table II also shows figures for a hybrid system, ΔPCM. This is differential pulse-code modulation[6] or, as described by van der Weg,[7] an *n*-digit delta modulation. Here the signal is sampled and quantized, as in PCM, but instead of each sample being coded in binary form, the differences between successive samples are coded. Such a system has an overload spectrum similar to ordinary delta modulation, and is therefore better suited for transmission of speech than plain PCM. The relation of ΔPCM to PCM and to delta modulation is illustrated in Fig. 4.

The new comparison is much more heartening. Delta modulation is still slightly behind PCM in quality for a given pulse rate, or in the pulse rate required for a given quality, but the difference is not very great, and would be negligible in many applications, if the equipment needed is really as simple as it appears. It is therefore necessary to take a closer look at the "black boxes" required for the two systems.

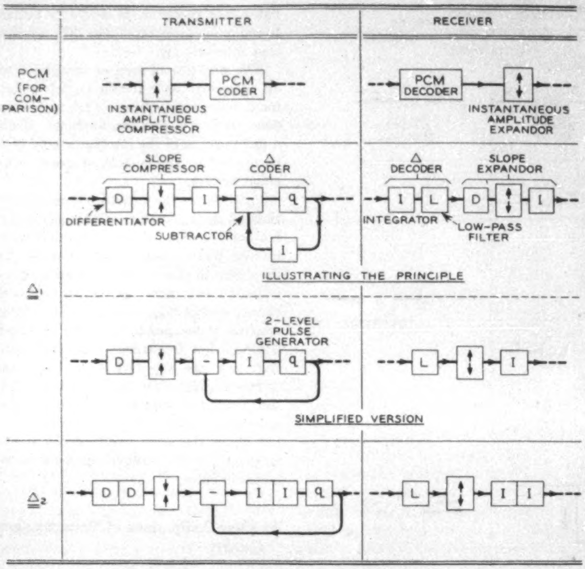

TRANSMITTER
RECEIVER
PCM (FOR COMPARISON)
PCM CODER
INSTANTANEOUS AMPLITUDE COMPRESSOR
PCM DECODER
INSTANTANEOUS AMPLITUDE EXPANDOR
SLOPE COMPRESSOR
Δ CODER
D
I
q
DIFFERENTIATOR
SUBTRACTOR
Δ DECODER
SLOPE EXPANDOR
L
INTEGRATOR
LOW-PASS FILTER
ILLUSTRATING THE PRINCIPLE
Δ₁
2-LEVEL PULSE GENERATOR
SIMPLIFIED VERSION
Δ₂

Fig. 5. Companding in delta modulation

Terminal Equipment

If our task is to send one speech channel by binary code, then there is not much room for argument. It has been shown how simple it is to make a delta coder and decoder: by comparison the PCM equipment is much more complicated. Both systems lend themselves to companding, so that strong signals can be compressed at the input, to prevent them from overloading the system and expanded again at the output. The type of compander suited to delta modulation differs a little from that of PCM, but its general effect and complexity are about the same. Syllabic or slow-acting companders may be used with any system. An instantaneous amplitude compressor, which flattens the peaks of large signals, works well for PCM, but is of no use for delta modulation, which overloads on slope and not on amplitude. Here a slope compressor is required. Such a device can be made out of an instantaneous amplitude compressor. Its use in delta modulation is shown in Fig. 5.

Delta modulation is the method of choice for coding speech channels one at a time, and if all communications in a telephone system are carried on in binary form, then there may well be a delta coder in every handset. This does not seem imminent, and at the moment the main applications in the telephone system for any binary transmission system are in: 1. transmitting wide-band signals, such as TV signals, and 2. sending many narrow channels simultaneously. How delta modulation fares in these fields will now be discussed.

On paper, there is no difference between coding a speech channel and coding a TV channel, and it is easy to draw up a set of boxes for each system, and compare them for complexity; see Fig. 6. At the top is the arrangement of a PCM terminal using a specially fabricated cathode-ray

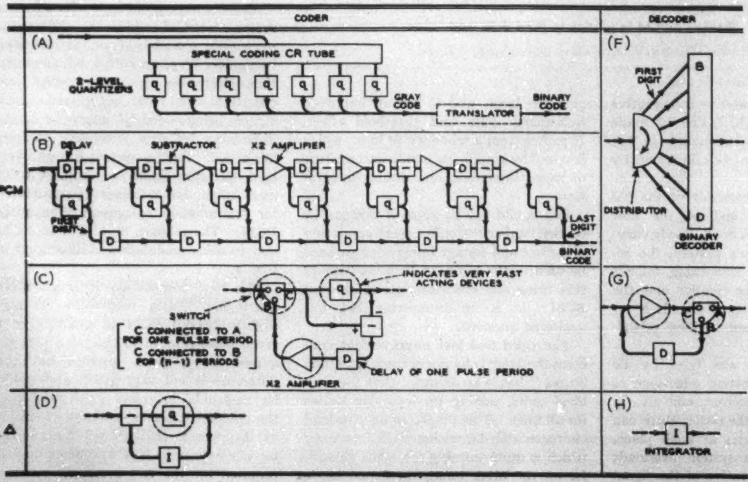

CODER
DECODER
(A)
SPECIAL CODING CR TUBE
2-LEVEL QUANTIZERS
q
GRAY CODE
TRANSLATOR
BINARY CODE
(B)
DELAY
SUBTRACTOR
X2 AMPLIFIER
PCM
FIRST DIGIT
D
LAST DIGIT
BINARY CODE
(C)
INDICATES VERY FAST ACTING DEVICES
SWITCH
C CONNECTED TO A FOR ONE PULSE-PERIOD
C CONNECTED TO B FOR (n−1) PERIODS
X2 AMPLIFIER
DELAY OF ONE PULSE PERIOD
(D)
Δ
I
(F)
FIRST DIGIT
B
DISTRIBUTOR
D
BINARY DECODER
(G)
(H)
INTEGRATOR

Fig. 6. Schemes for encoding a single wide channel

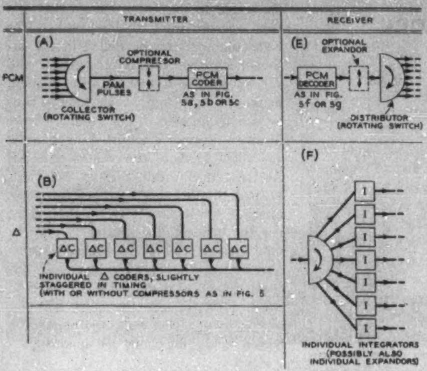

TRANSMITTER
RECEIVER
PCM
(A)
OPTIONAL COMPRESSOR
PAM PULSES
PCM CODER
AS IN FIG. 5a, 5b OR 5c
COLLECTOR (ROTATING SWITCH)
(E)
OPTIONAL EXPANDOR
PCM DECODER
AS IN FIG. 5f OR 5g
DISTRIBUTOR (ROTATING SWITCH)
Δ
(B)
ΔC
INDIVIDUAL Δ CODERS, SLIGHTLY STAGGERED IN TIMING (WITH OR WITHOUT COMPRESSORS AS IN FIG. 5
(F)
I
INDIVIDUAL INTEGRATORS (POSSIBLY ALSO INDIVIDUAL EXPANDORS)

Fig. 7 (left). Schemes for encoding many narrow channels

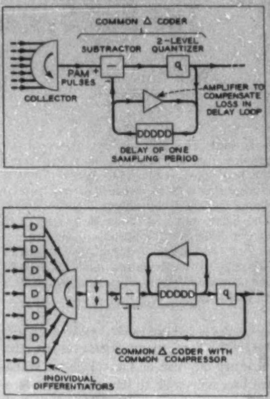

COMMON Δ CODER
SUBTRACTOR
2-LEVEL QUANTIZER
PAM PULSES
COLLECTOR
q
AMPLIFIER TO COMPENSATE LOSS IN DELAY LOOP
DDDDD
DELAY OF ONE SAMPLING PERIOD
D
DDDDD
q
COMMON Δ CODER WITH COMMON COMPRESSOR
INDIVIDUAL DIFFERENTIATORS

Fig. 8 (right). A many-channel delta coder

coding tube. This is followed by two versions of coders that have been worked on in the Bell Telephone Laboratories, which use only conventional elements. Finally, in all its elegance, is the delta modulation equipment.

The only snag is that in the present state of the art, it is very difficult to build some of the boxes shown in this figure. To send commercial TV over delta modulation, a pulse rate of around 100 megacycles is needed. For the circuit in Fig. 6(D) therefore, a pulse generator is needed (the box *q* circled in the diagram) which will make the decision of whether to send a pulse or not, then will generate a pulse, and then recover to make a new and independent decision, all in the space of 1/100th of a microsecond. With present tubes and transistors, it is anticipated that boxes may be constructed which make 10 million decisions per second, but it would be exceedingly difficult to make them ten times as fast. (There are, it is true, other ways of producing delta modulation which will tolerate slower decision elements but these become very unwieldy; they are no longer much simpler than PCM coders.) At the moment, then, if TV is sent in binary pulse form, PCM is the logical choice, probably using arrangements A or B of Fig. 6.

There are several ways of sending many channels simultaneously in binary code, one of which would be to combine all the signals first, using frequency-multiplexing. The composite signal is then treated like a single wide-band channel. This is, however, far too elaborate and costly. The simplest way of coding many channels simultaneously is to use time division.

Here one coder only is required, and it looks in turn at the amplitudes of all the incoming channels. See Fig. 7 (A) for a PCM system. Travelling along the line, one finds a group of pulses belonging to channel 1, followed by a group of channel 2, etc. At the decoder, each group is treated to give back a sampled amplitude, and this is then routed to the right channel. In such a system, the cost of coder and decoder can be shared by many channels, and can become quite insignificant.

Delta modulation does not lend itself so easily to common coding and decoding, and one would normally use the arrangement shown in Fig. 7 (B). No matter how simple the delta coders are, the arrangement of one to a channel is likely to be extravagant. With some ingenuity one can actually devise a delta coder that is common to several channels, and even one that includes common companding, see Fig. 8, but this is now no longer a simple circuit. Furthermore, the collecting and distribution switches for delta modulation must work very much faster than those of PCM, perhaps five to ten times as fast. This makes troubles from crosstalk more severe, and the switches more complicated and expensive.

Conclusions

What then is the use of delta modulation to the transmission engineer? It has been seen that there are several advantages in transmitting signals in binary code. This is an exciting development, which is likely to become more important as systems get more complicated and repeaters get smaller and cheaper.

Delta modulation is the simplest and cheapest known of converting a continuous signal into code form, and is to be preferred whenever coding one channel at a time is desired. It is not quite as efficient as its rival, PCM, in getting the maximum quality out of a given pulse rate, but when suitably doctored, it does not lag far behind.

If the signal to be transmitted is of wide bandwidth, delta modulation reaches its limitations some time before PCM, and in the present state of the art it would be exceedingly hard to make a delta system fast enough for commercial TV. Finally, delta modulation does not lend itself as readily to common equipment use as does PCM, and is not a good system where it is desired to code and decode many channels simultaneously. Delta modulation will probably always have some place in the world of communications but it is likely to be a more modest place than that of PCM.

References

1. A Comparison Between Delta and Pulse Code Modulation, L. H. Zetterberg. *Ericsson Technics*, Stockholm, Sweden, vol. 2, 1955, pp. 95–154.

2. Deltamodulation, F. de Jager. *Philips Research Reports*, Eindhoven, Netherlands, vol. 7, 1952, pp. 442–66.

3. The Philosophy of PCM, B. M. Oliver, J. R. Pierce, C. E. Shannon. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 36, 1948, pp. 1324–31.

4. Experimental Multichannel Pulse Code Modulation System of Toll Quality, L. A. Meacham, E. Peterson. *Bell System Technical Journal*, New York, N. Y., vol. 27, 1948, pp. 1–43.

5. Spectra of Quantized Signals, W. R. Bennett. *Ibid.*, vol. 27, 1948, pp. 446–72.

6. Differential Quantization of Communication Signals, C. C. Cutler. *U. S. Patent No. 2,605,361*, July 29, 1952.

7. Quantizing Noise of a Single Integration Delta Modulation System with an N-Digit Code, H. van der Weg. *Philips Research Reports*, Eindhoven, Netherlands, vol. 8, 1953, pp. 367–85.

The Bell System Speakerphone

W. F. CLEMENCY
NONMEMBER AIEE

F. F. ROMANOW
NONMEMBER AIEE

A. F. ROSE
FELLOW AIEE

Synopsis: The Bell System Speakerphone provides hands-free telephony. It permits the telephone customer to receive over a loudspeaker and to transmit over a microphone located a convenient distance away. The techniques and components employed in the design of the Speakerphone are described. A survey of customer reactions to this type of telephone set indicates that 90 per cent are fairly or entirely satisfied with its performance.

THERE are many instances in the every-day use of the telephone when the user may wish to have his hands free while answering a call. The new Bell System Speakerphone introduced recently for general use provides this feature. The user now talks into a microphone which can be placed at arm's length and listens to a small loudspeaker placed at the same distance from the person but separated from the microphone by about 3 feet. He can dial without holding the handset and, during the telephone conversation, write notes, or refer to any reference material. It is also possible for a small group to participate in the telephone conversation. Further, a physically handicapped person, once a call is set up, can converse freely without the need to hold the handset. To allow for privacy when needed and to insure telephone service in the event of power-line failure, the Speakerphone is equipped with a handset. A call is transferred to the handset merely by removing it from the cradle.

There are limitations to the use of the Speakerphone. In certain rooms, where there is little sound absorption, the resulting reverberation may interfere with the reception at both ends of the telephone connection. In particular the person at the distant end when listening may note a quality such as is observed when speaking into an empty barrel and when talking, an enhanced sidetone or echo of his own voice. Another serious disturbing factor is ambient noise at the Speakerphone location, which in combination with a low incoming signal may cause the customer to revert to the handset. The general question of such disturbances is treated in a paper by Emling.[1] Field experience, however, has shown that conditions of this kind are less likely to be encountered in locations where Speakerphone installations have been made after checking the attendant circumstances and usage. For some time the need for quiet and nonreverberative surroundings in business offices has resulted in the extensive use of absorptive materials for ceiling and wall treatment. Under such conditions, the Speakerphone provides generally acceptable service.

General

At present, two types of Speakerphone are available, the one shown in Fig. 1, coded the *595*-type telephone set, is primarily an individual line set, while that in Fig. 2, the *1A* Speakerphone system, can be used for the same purpose as well as for multiline key telephone service. In the first type, the housing is similar in appearance to the *500*-type telephone set,[2] but it contains in addition to the usual components a microphone, ON and OFF keys, and a loudspeaker volume control as illustrated in Fig. 3. It will be noted that the microphone is placed at the front of the set. This location is closest to the user, and he will tend to remain within arm's distance because of the need to operate the dial and the controls. It is desirable that this distance be minimized to reduce the effects of room reverberation on the transmitted speech, and to permit the use of less amplification in the transmit circuit so that more amplification is available in the receive circuit as will be explained later. The second type has the microphone, the ON and OFF keys, and the volume control in a separate small unit. This permits greater freedom in placing the microphone and, if located closer to the user, results in a transmission improvement. In both types, the loudspeaker is mounted in the small plastic housing shown in the upper center of Fig. 1 and 2.

The Speakerphone is easy to use. To place a call, it is only necessary to press momentarily the ON key which lights up as an indication that the set is connected to the line. Then, when dial tone is heard from the loudspeaker, the called number can be dialed in the ordinary manner. When answering a call, the ON button is momentarily pressed, again causing it to light up, and the conversation can start. The volume of the incoming speech is adjusted by turning the small knob of the volume control. At the end of a call, the OFF button is pressed to disconnect the set from the line. If, during a call, it is desired to use the handset, the user need only to life it and the call is automatically transferred from the Speakerphone. If the handset is being used, and it is desired to transfer to the Speakerphone, the ON button is depressed while the handset is placed on the telephone set mounting.

Components

An interior view of the telephone set is shown in Fig. 4. The microphone and an associated equalizer, the controls, and a set of terminals are mounted on a common bracket which is fastened to the base. The dial of the set is equipped with two sets of contacts which close when dialing; one set prevents dial clicks in the handset while the other prevents such clicks in the loudspeaker.

In addition to the units placed on the desk, an apparatus box, Fig. 5 contains the receive and transmit amplifiers, power supply, hybrid coil, and a relay used for the switching functions. An interior view is shown in Fig. 6. The apparatus box is connected with the other components by a suitable length of cabling so that it may be mounted in an inconspicuous location. The power is obtained from a 120-volt 60-cps (cycle-per-second) source and the consumption is 0.6 watt in the stand-by and 8.0 watts in the usage condition. All amplifier components except power supply and transformer are mounted on a printed circuit card which plugs into a connector on the apparatus box in order to facilitate maintenance. The printed-circuit amplifiers, one for the transmitter and one for the loudspeaker, are identical in design and use filamentary-type tubes which warm to operating temperature in about 0.2 second. A potentiometer to regulate gain is a part of each amplifier unit and is preset at the factory.

Design Features

The basic circuit of the Speakerphone is shown in Fig. 7. The hybrid coil, which is a form of Wheatstone bridge,

---

Paper 57-139, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted November 28, 1956; made available for printing December 3, 1956.

W. F. CLEMENCY and F. F. ROMANOW are with Bell Telephone Laboratories, Inc., Murray Hill, N. J., and A. F. ROSE is with the American Telephone and Telegraph Company, New York, N. Y.

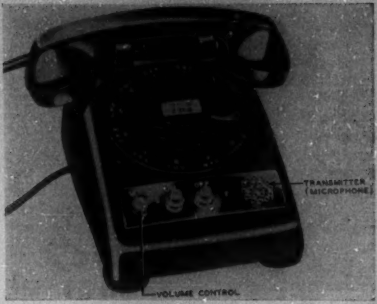


Fig. 1 (left). 595-type Speakerphone

Fig. 3 (above). 595-type Speakerphone housing of telephone set

couples the 2-way telephone line to the output of the 1-way transmitting amplifier and to the input of the 1-way receiving amplifier, and introduces a loss, L_H, between them. As indicated in Fig. 7, the unavoidable acoustic coupling, L_A, through the air path between the loudspeaker and the transmitter completes a feedback loop. When there is a net power gain at any frequency in the transmission range in this closed loop, it acts as an oscillator and singing will occur. An upper limit is therefore placed on the sum of the gain, G_T, of the transmitting amplifier and the gain, G_R, of the receiving amplifier. Expressing the gains and losses in decibels, singing does not occur when the following condition is satisfied:

$$|G_T|+|G_R|+|L_H|+|L_A|\leq 0$$

Since the acoustic loss, L_A, is a highly variable factor and is dependent on the individual reflections from the surfaces of the room and ambient objects, no control if its phase is possible. Hence the magnitude of the quantities are chosen for the foregoing expression.

The hybrid coil loss, L_H, is determined by the degree of impedance balance between the telephone and the balancing network. If a perfect balance could be obtained at all frequencies, this loss would be infinite, and singing could not occur. However, the impedance of telephone lines varies considerably with frequency in magnitude and phase. It may be resistive, capacitive, or inductive, depending on the length and makeup of the loop to the central office, and on the length and type of trunk between central offices. The variability of impedances is especially true for private branch exchange installations, where internal calls may be on a short loop and external calls on a long loop. There is, fortunately, a rough correlation between loop impedance and loop direct current, and advantage is taken of this fact in the design of the balancing network. The operation is similar to that of the variable anti-sidetone network of the *500*-type telephone set.[8] The action is as follows: The d-c path through the Speakerphone circuit is from one side of the telephone line through a winding of the hybrid coil to the inductance coil, *L*, and thence to the other side of the telephone line. The voltage drop across the inductance coil, which has an 85-ohm resistance, serves to regulate the impedance of the variable network. The variable component is a symmetrical silicon carbide varistor, the resistance of which is high for low direct voltages and low for high direct voltages. Fig. 8 indicates the relationship between the loop direct current and the impedance of the balancing network. The increase that this network offers in the permitted receiving gain before singing occurs over that with a nonreactive 1,000-ohm resistance is shown in Fig. 9. The plotted points above the dashed line represent increases and the points below represent decreases. These data were obtained on 30 telephone connections representing a range of battery, trunk, and loop conditions most likely to be encountered in service. It will be noted that the reductions occur on some short-loop connections where the amount of receiving gain required is small and

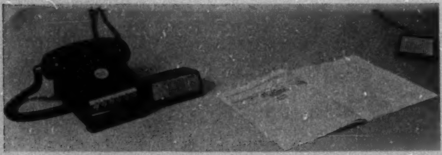


Fig. 2 (above). Speakerphone system no. 1A

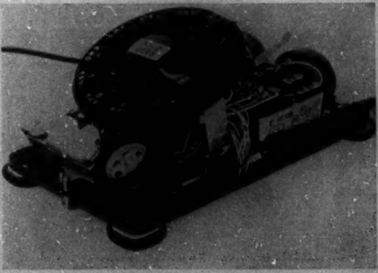


Fig. 4 (right). 595-type Speakerphone, interior of telephone set

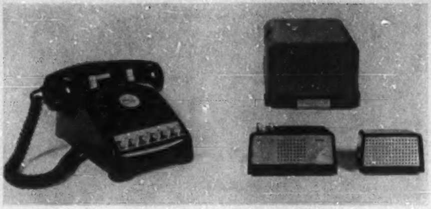


Fig. 5. Speakerphone system no. 1A components

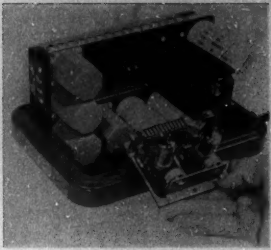


Fig. 6 (right). Apparatus box and printed circuit amplifiers

singing is not a limiting factor. On the average, the variable network shows a 5 db (decibels) better balance than a fixed 1,000-ohm resistance.

The incoming signal from the telephone line is supplied to the receiving amplifier through the hybrid coil, a 2-microfarad blocking capacitor, and a volume control which consists of two potentiometers with a common shaft arranged in a balanced circuit to minimize crosstalk and electrical noise pickup. A fixed resistor, between the potentiometers, assures that the gain cannot be reduced below a minimum sufficient to enable the loudspeaker to produce dial tone at an easily heard level. In addition, most local calls may be received with the volume control at this minimum position.

An especially designed thermistor-resistance combination is shunted across the output of the receiving amplifier, as shown in Fig. 7. By limiting the steady-state output voltage of the amplifier, the thermistor assures that the singing tone, when it occurs, cannot exceed 70 db sound level[4] at the normal listening distance. Thus, a disturbing howl is reduced to a soft audible tone. The limiting action of the thermistor also takes place in the case of speech. Because of the fluctuations in this type of signal and the integrating properties of the thermistor, speech pressures may reach an 80-db sound level. This sound pressure is approximately 10 db above the level of conversational speech at one meter and is sufficient to provide satisfactory reception for noise levels below 60 db. The volume-limiting action of the thermistor is also valuable in reducing talker echo at a connected telephone. This effect results when the received loudspeaker signal is high and occurs primarily on local connections.

The amplifiers have two stages of conventional class *A* design employing input and output transformers with resistance-capacitance interstage coupling. The transmitting amplifier is adjusted at the factory to have an "ideal transducer" gain of 51 db at 1,000 cps. For a person speaking 18 inches from the microphone this setting produces a line level which is approximately equal to that obtained from a *302*-type telephone with the same speaker 1 inch away from the handset transmitter. A decreasing high-frequency response is provided to help equalize the microphone characteristic. The receiving amplifier is adjusted to a 44-db gain, which can be reduced 33 db by the manual volume control. A decreasing high-frequency response serves to balance the rising characteristic of the loudspeaker.

Acoustical Considerations

The microphone which is employed in this set is a modified *U1* receiver[5] unit enclosed in a small case to give it a definite acoustic chamber volume. To minimize irregularities in the transmit response resulting from reflections of sound by the desk surface on which the Speakerphone rests, the microphone is mounted in front of the set and as close to this surface as possible. The microphone mounted in this way gives a response which rises in the high-frequency range. To compensate for this an equalizer is connected across the unit and the resulting response, including the change introduced by the amplifier characteristic, is shown in Fig. 10(A). By experiment, this response has been found to be a good compromise for speech quality, for the reduction of the effects of room reverberation, for the reduction of talker echo, and for the control of the singing characteristics of the set. Basically the microphone is a nondirectional pressure unit, but because of the size of the telephone set, it has some directivity at frequencies above 1,000 cps. However, this directivity is small enough so that the

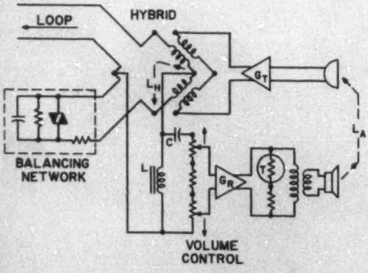


Fig. 7. Basic circuit of Speakerphone

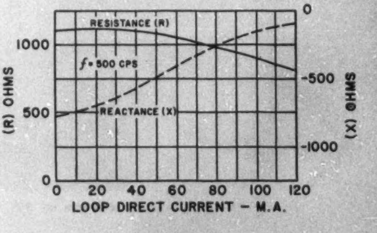


Fig. 8. Impedance of balancing network for different loop direct current

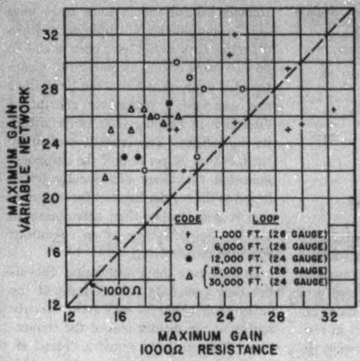
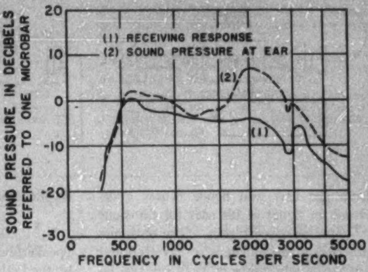


Fig. 9 (left). Maximum gain, variable network compared with 1,000-ohm resistance

Fig. 11 (right). Receive response, effect of head diffraction



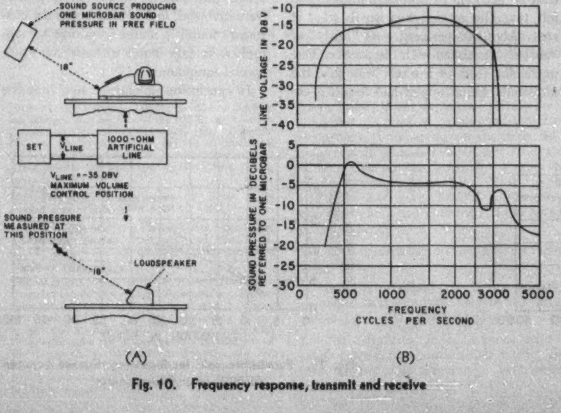
total variation in transmit level for speech originating equidistant from the microphone is only about 5 db for all directions around the set.

The receive response of the Speakerphone is shown in Fig. 10(B). The decreasing response at the higher frequencies produces a more pleasant quality for speech received from a connected handset telephone over local lines without producing a serious degradation of articulation on long lines. This results because the higher frequencies are augmented by the sound diffraction effects of the listener's head[6] so that the sound pressure at the entrance of the ear canal actually rises at frequencies between 1,500 and 3,000 cps; see Fig. 11. These reflections do not occur with the regular telephone receiver held close to the ear. The loudspeaker is also nondirectional at the lower frequencies and slightly directional because of its size at the higher frequencies. A small loudspeaker was selected to minimize the size of the desk-top components and to provide flexibility in its placement.

It has been found that better control of the singing properties of the set is obtained when the transmit frequency response is approximately complementary to the receive response. In this case, the singing loop has a relatively flat frequency characteristic which permits maximum effective transmit and receive gain. However, a set having this characteristic is likely to sing at any frequency. In order to prevent singing at frequencies above 1,000 cycles, easily produced by bringing the hand or any small object near the microphone, it is better to tailor the loop characteristic to sing at low frequencies.

A method of studying the singing loop is outlined in Fig. 12. In this method, the loudspeaker is disconnected from the set and a 10-ohm resistance which is equal to the nominal impedance of the speaker is connected in its place. An oscillator maintains a constant voltage, E_1, across the loudspeaker. The voltage, E_2, returned to the 10-ohm resistance, is measured. The set is arranged on a table top in the recommended operating position. The voltage, E_1, applied to the loudspeaker is 0.1 volt which is within the range applied to the loudspeaker during normal operation of the set. The ratio of E_2 to E_1 is a transmission measurement of the singing loop and is called the return voltage ratio. The set will sing when this voltage ratio is greater than one. Adjustment of the volume control to a lower gain position reduces this ratio and singing will cease when the ratio becomes less than one. To obtain reproducible conditions, the variable balancing network is replaced with a 1,000-ohm resistor and the line terminals of the set are connected to 800 ohms. This produces an unbalance in the hybrid coil of a definite, easily reproduced amount. A measurement of this return ratio is shown in Fig. 13. These data indicate that the return ratio has a fairly flat frequency characteristic with no sharp peaks and they also indicate that singing would occur at a low frequency, about 600 cps, which is desirable.

The other factor influencing singing is the acoustic coupling between the microphone and the loudspeaker. With nondirectional transducers, this coupling is controlled by the distance between the instruments, becoming less as the distance is increased. The variation in permissible gain as the distance between microphone and loudspeaker is changed is shown in Fig. 14. An increase in this distance allows more gain to be used in the loudspeaker circuit. However, beyond a certain separation, further increase in the distance between the instruments moves the loudspeaker further



(A)

(B)

Fig. 10. Frequency response, transmit and receive

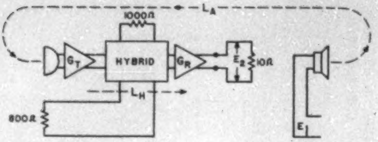


Fig. 12. Circuit for singing-loop study

from the user and hence causes a decrease in signal at his ears for the same loudspeaker circuit gain. The optimum practical arrangement is to have the user approximately midway between the instruments which are separated by about 3 feet on the desk top. Fig. 14 also indicates that the acoustic properties of the room influence the permissible gain which may be used in the loud speaker circuit. In a reverberant room having little sound absorption, the permissible gain before singing occurs is less than that in a room with good sound absorption.

In experimental sets of this type, reduction of acoustic coupling between the loudspeaker and the transmitter was attempted by using various types of directional instruments for the microphone and also for the loudspeaker. It was found that low coupling could be obtained in a free-field room with some combination of directional instruments. However, this coupling increased considerably in a room having the acoustic properties of normal offices. Also, this coupling was influenced considerably by the position of ambient objects and persons in the vicinity of the loudspeaker and the microphone. The reduction in the acoustic coupling obtained by increasing the distance between the microphone and the loudspeaker seems to be less adversely influenced by room acoustics and ambient objects than the reduction obtained through transducer directivity. The use of nondirectional transducers also permits greater flexibility in design of housings and the placement of the transducers.

Application in the Bell System

Up to the present time, about 75,000 Speakerphones have been put into service by the Bell System companies throughout the United States. They are used in a great variety of situations but more extensively in business offices than in residences. Moreover, there is a wide diversity in the type of organizations that have found them of value.

In view of the fact that the Speakerphone is subject to transmission limitations, and is now being used in sufficient numbers that it might have an over-all impact on customer reaction to telephone service in general, a survey was made on the consumer level. This was directed at determining the degree of satisfaction experienced by the users of the Speakerphone, and whether they felt that it was proving to be worth its cost to them. Interviews were conducted on the premises of some 500 customers and comparable questions were asked of all. The sample selected covered all types of customers, including heavy and light industries, welfare organizations, financial institutions, government agencies, and utilities.

It does not seem necessary to include here all the details of the questions and the extensive analysis made of the answers. It probably is sufficient to say that 85 per cent of the users interviewed felt that the service was worth what it cost. About 90 per cent were "fairly" or "entirely" satisfied with the service. One particular item of interest was that the physically handicapped had found it extremely helpful and in some cases were able to continue useful employment when otherwise they would have been unable to do so. Most of the customers were aware that the equipment had some limitations and exercised a reasonable amount of discretion in keeping their usage within its capabilities. This is made evident by the fact that 25 per cent of the customers indicated that they occasionally transferred to the regular handset for transmission reasons rather than attempting to use the Speakerphone when conditions appeared too adverse.

In analyzing the transmission difficulties that were encountered, most of the unfavorable reactions were from the subscriber at the distant end of the connection. Frequently he noted a "barrel effect" and commented unfavorably on this to the user of the Speakerphone. The Speakerphone customer himself occasionally had trouble with inadequate volume on connections involving a low-volume talker at the distant end of the circuit, or on some switched connections which may have had more than average transmission loss. In some of these cases, he may have attempted to correct the situation by turning his volume control too high and thus produced howling which was noted by the party at the distant end.

The primary purpose of the Speakerphone is to leave the hands free during telephone conversations. In addition to this primary purpose, however, several of the customers noted other advantages. For example, while waiting for the distant party to answer they need not hold the receiver to their ear, and a few mentioned the additional comfort and avoidance of fatigue in the case of calls lasting for an extended period. There were a number of favorable comments on group usage particularly when small groups were involved, and in some cases the customers found it useful to permit stenographers to take notes without any other special equipment.

In conclusion, it may be said that the

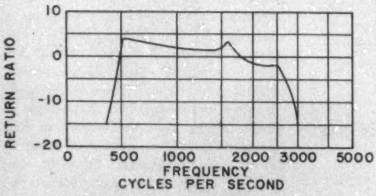


Fig. 13. Singing-loop return ratio

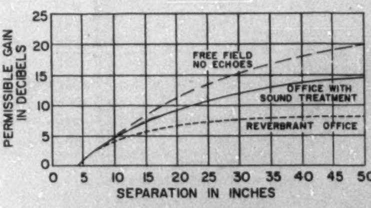


Fig. 14. Permissible gain for increasing distance between microphone and loudspeaker

present type of Speakerphone fills a real need and its use will undoubtedly continue to expand. If some of the technical difficulties mentioned in another paper[1] and noted by the users can be overcome, an even larger field of application appears to be indicated.

References

1. GENERAL ASPECTS OF HANDS-FREE TELEPHONY, J. W. Emling. *AIEE Transactions*, pt. I, May 1957, pp. 201–05.
2. AN IMPROVED TELEPHONE SET, A. H. Inglis, W. L. Tuffnell. *Bell System Technical Journal*, New York, N. Y., vol. 30, Apr. 1951, pp. 239–70.
3. AN IMPROVED CIRCUIT FOR THE TELEPHONE SET, A. F. Bennett. *Ibid.*, vol. 32, May 1953, pp. 611–26.
4. ACOUSTICAL TERMINOLOGY. *ASA Standard Z241*, American Standards Association, New York, N. Y., 1951.
5. THE RING ARMATURE TELEPHONE RECEIVER, E. E. Mott, R. C. Miner. *Bell System Technical Journal*, vol. 30, Jan. 1951, pp. 110–40.
6. PRESSURE DISTRIBUTION IN THE AUDITORY CANAL, F. M. Wiener, D. A. Ross. *Journal*, Acoustical Society of America, New York, N. Y., vol. 18, 1946, pp. 401–08.

Analog Computer Representation of Alternators for Parallel Operations

LUIZ V. BOFFI
NONMEMBER AIEE

VINTON B. HAAS, JR.
ASSOCIATE MEMBER AIEE

THE USE of dynamically faster drives in aircraft electric-power-generating systems precludes the assumption formerly made in computer studies of these systems that the electrical transient is much faster than the mechanical transient. Furthermore, the representation of alternators for parallel operation introduces new conceptual problems not formerly encountered in representing the alternator for isolated operation. In this paper, the basic circuit equations developed by Park are used to describe the alternator. An additional form is also used so that the effect of saturation is approximated. The usual procedure of breaking the system into its simplest units for system studies is followed in this presentation. The representations of the constraints imposed by the interconnection of the units are separated from the units themselves.

The paper deals with the conceptual problems of adapting the basic equations for a single alternator with an approximate saturation correction to the condition of parallel operation. In addition some unique problems, encountered in placing the equations on an analog computer, are discussed. The representation given in this paper has been used during an analog computer study of an aircraft electric-power-generating system, involving operation of alternators in parallel. The representation given here is suitable for studying the paralleling of isolated alternators and for the application and removal of balanced loads on isolated or paralleled systems.

The Alternator Equations

The basic circuit equations[1,2] of the alternator in terms of the 2-axis variables are dependent on the following assumptions:

1. Smooth stator and pole faces. This is equivalent to neglecting slot effects.
2. Eddy current and hysteresis effects are neglected.
3. Saturation is neglected. (The form of the computer representation is, however, adapted so as to include an approximate representation of saturation.)
4. Sinusoidally distributed magnetomotive forces (mmf) in the armature. This allows the armature mmf to be resolved into two components; one acting on the main field pole, the direct-axis component; and the other acting on the interpolar space, the quadrature-axis component. This resolution leads to the adoption of fictitious armature currents i_d and i_q instead of the actual phase currents.
5. No zero-sequence component of current is required.
6. Damper windings are represented by two equivalent windings, one in the direct axis and one in the quadrature axis.
7. For simplicity, it is assumed that a single mutual inductance exists among the three circuits in the direct axis: field, armature, and damper.

Assumptions 1–4 and 7, are pertinent to the equations presented in Concordia's book,[1] while assumptions 5 and 6 are introduced in order to simplify the problem and may be relinquished very easily. Assumption 5 restricts the representation to the case of nonconnected neutrals.

The equations for an alternator with the foregoing assumptions written in a per-unit system are as follows:

$$v_f=r_f i_f+\left(\frac{p}{\omega_0}\right)(\omega_0\psi_f) \tag{1}$$

$$v_d=-r_a i_d+\left(\frac{p}{\omega_0}\right)(\omega_0\psi_d)-\left(\frac{w}{\omega_0}\right)(\omega_0\psi_q) \tag{2}$$

$$v_q=-r_a i_q+\left(\frac{p}{\omega_0}\right)(\omega_0\psi_q)+\left(\frac{\omega}{\omega_0}\right)(\omega_0\psi_d) \tag{3}$$

$$0=r_{kd} i_{kd}+\left(\frac{p}{\omega_0}\right)(\omega_0\psi_{kd}) \tag{4}$$

$$0=r_{kq} i_{kq}+\left(\frac{p}{\omega_0}\right)(\omega_0\psi_{kq}) \tag{5}$$

where

v = voltage
i = current
ψ = flux linkage
r = resistance
ω = instantaneous angular speed
ω_0 = rated angular speed
p = time differential operator d/dt
$X=\omega_0 L$, the reactance at base frequency

and the subscripts

f = main field
d = direct axis
q = quadrature axis
a = armature
k = damper

The flux linkages are given by

$$(\omega_o\psi_f)=(\omega_0\psi_{md})+X_{fl}i_f \tag{6}$$

$$(\omega_0\psi_d)=(\omega_0\psi_{md})-X_{al}i_d \tag{7}$$

$$(\omega_0\psi_q)=(\omega_0\psi_{mq})-X_{al}i_q \tag{8}$$

$$(\omega_0\psi_{kd})=(\omega_0\psi_{md})+X_{kdl}i_{kd} \tag{9}$$

$$(\omega_0\psi_{kq})=(\omega_0\psi_{mq})+X_{kql}i_{kq} \tag{10}$$

where the first index *m* refers to a mutual and the second index *l* refers to a leakage.

The mutual flux linkages may be expressed by:

$$(\omega_0\psi_{md})=X_{md}(i_f-i_d+i_{kd}) \tag{11}$$

$$(\omega_0\psi_{mq})=X_{mq}(-i_q+i_{kq}) \tag{12}$$

Paper 57-150, recommended by the AIEE Computing Devices Committee and the Air Transportation Committee of the Applications and Industry Division, and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957.

Manuscript submitted October 15, 1956; made available for printing December 11, 1956.

LUIZ V. BOFFI is with Massachusetts Institute of Technology, Cambridge, Mass., and VINTON B. HAAS, JR., is with The University of Connecticut, Storrs, Conn.

This work was conducted by the Energy Conversion Group of the Servomechanisms Laboratory, Massachusetts Institute of Technology, and was supported, in part, by Wright Air Development Center under *Contract AF33(616)3242*. The authors wish to acknowledge the assistance of the members of the Energy Conversion Group and in particular Dr. M. Riaz and R. Nease. R. Harnett of the WADC analog computer facility also aided in the work during the computer studies.

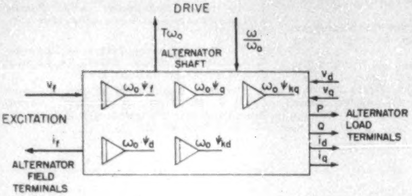


Fig. 1. Block representation of the alternator

These equations are the result of a choice of per-unit system in which the mutual inductances are the same for both stator and rotor. This choice of per unit is particularly convenient in scaling variables for application on an analog computer.

Thus, for the alternator operating at no load, rated speed, and rated terminal voltage in the steady state according to equation 2

$$v_d = -\omega_0\psi_q = 0$$

and according to equation 3

$$v_q = +\omega_o\psi_d = 1.0$$

From equations 7 and 11

$$1.0 = (\omega_0\psi_d) = X_{md}i_f$$

Therefore, the base field current is the value which produces an air-gap voltage equal to X_{md} on the air-gap line at rated speed. The base armature current and voltage are the rated rms values.

The active and reactive power developed by the alternator are

$$P = v_di_d + v_qi_q \tag{13}$$

$$Q = -v_di_q + v_qi_d \tag{14}$$

where the base P and Q is the alternator rated volt-amperes.

The electromagnetic torque developed by the alternator is

$$(T\omega_0) = (\omega_0\psi_d)i_q - (\omega_o\psi_q)i_d \tag{15}$$

The base value of torque is that torque which produces rated power at rated speed, and the torque in newton-meters will be equal to $(T\omega_o)$ times the base value of torque. Torque in the per-unit system used in this paper is $(T\omega_o)$.

Representation of the Equations on an Analog Computer Neglecting Saturation

In the adaptation of the equations to an analog computer, it is desirable to avoid differentiation. Equations 1 through 5 are thus reformed as follows:

$$\omega_0\psi_f = \frac{\omega_0}{p}[v_f - r_fi_f] \tag{1A}$$

$$\omega_0\psi_d = \frac{\omega_0}{p}\left[v_d + r_ai_d + \frac{\omega}{\omega_0}(\omega_o\psi_q)\right] \tag{2A}$$

$$\omega_o\psi_q = \frac{\omega_o}{p}\left[v_q + r_ai_q - \frac{\omega}{\omega_0}(\omega_0\psi_d)\right] \tag{3A}$$

$$\omega_0\psi_{kd} = \frac{\omega_0}{p}[-r_{kd}i_{kd}] \tag{4A}$$

$$\omega_0\psi_{kq} = \frac{\omega_0}{p}[-r_{kq}i_{kq}] \tag{5A}$$

These equations indicate the five integrations that are necessary for the computer representation. The five values of $\omega_0\psi$ generated by these equations may be used in equations 6 through 12 to generate the five corresponding values of current and the $(\omega_0\psi_{md})$ and $(\omega_0\psi_{mq})$. Referring again to equations 1, 2, and 3, it is seen that the signals v_f, v_d, v_q, and ω/ω_0 are supplied to the representation of the alternator to complete the solution of the equations.

In addition to the five integration operations, the following multiplications must also be performed:

In equation 2

$$\frac{\omega}{\omega_0}(\omega_0\psi_q)$$

In equation 3

$$\frac{\omega}{\omega_0}(\omega_0\psi_d)$$

In equation 13

v_di_d and v_qi_q

In equation 14

v_di_q and v_qi_d

In equation 15

$(\omega_0\psi_q)i_d$ and $(\omega_0\psi_d)i_q$

The final result is a representation for the alternator as a unit as shown in Fig. 1. This unit corresponds to the physical apparatus at the field terminals, the shaft, and the load terminals, and may be interconnected accordingly with representations of the drive, the excitation system, or the load. In this form, the representation is most convenient since from the terminals the operator may consider the representation as analogous to the hardware. Where the signals are in per unit it is a simple matter to choose the proper scale factors for application to a computer as is discussed earlier.

Representation of the Saturation Nonlinearity

Saturation in the alternator results in a nonlinear relation between flux and mmf. In general, this nonlinear effect in salient-pole machines should be introduced in a different manner in the armature and poles. In the armature, the resultant flux dictates the saturation level and affects this flux and the corresponding mmf. In the poles, the saturation level is fixed by a combination of d axis and field leakage flux and affects the relation between the d axis flux and the d axis mmf.

To introduce all these effects would entail the introduction of two different nonlinearities with one of them depending on two variables. Considerably more data than commonly available (open-circuit characteristic) would be required. In substitution, the rather crude and much simpler approximation of considering all the leakage reactances and the mutual reactance X_{mq} independent of saturation is used. Thus, only the direct-axis mutual reactance, X_{md}, is considered to change with saturation. X_{md} appears in equation 11 only.

With the alternator at no load, constant speed, the terminal phase voltage

$$v_t = v_q = \omega_0\psi_{md} = X_{md}i_f \tag{16}$$

thus the nonlinear form for X_{md} is determined from the open-circuit magnetiza-

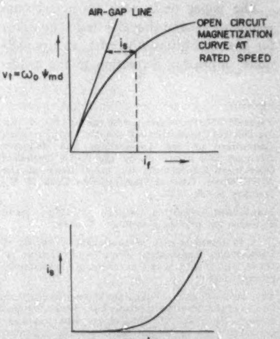


Fig. 2. Obtaining the $i_s = f(\omega_o\psi_{md})$ from the open-circuit characteristic

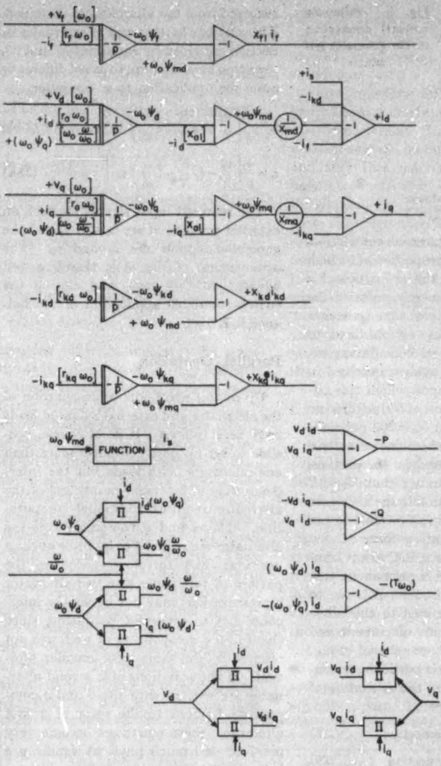

+vf [ωo]
-if
[rf ωo]
-1/p
-ωo ψf
-1
xf1 if
+ωo ψmd
+vd [ωo]
+id
[ra ωo]
+(ωo ψq)
[ωo ω/ωo]
-ωo ψd
-id
[Xal]
+ωoψmd
1/Xmd
+is
-ikd
-if
+id
+vq [ωo]
+iq
-(ωo ψd)
-ωo ψq
-iq
+ωoψmq
1/Xmq
-ikq
+ iq
-ikd
[rkd ωo]
-ωo ψkd
+xkd ikd
+ ωo ψmd
-ikq
[rkq ωo]
-ωo ψkq
+xkq ikq
+ωo ψmq
ωo ψmd
FUNCTION
is
vd id
vq iq
-P
-vd iq
vq id
-Q
(ωo ψd) iq
-(ωo ψq) id
-(Tωo)
id
iq
Π
id(ωo ψq)
ωo ψq
ωoψq ω/ωo
ω/ωo
ωo ψd ω/ωo
ωo ψd
iq (ωo ψd)
vd
vd id
vq iq
vq id
vq

Fig. 3. Computer diagram for the alternator

tion curve of the alternator according to equation 16, and equation 11 is represented on the computer by

$$\omega_0\psi_{md} = f(i_f - i_d + i_{kd}) \tag{11A}$$

where f is the nonlinear X_{md} and depends on the magnitude of the signal $(i_f - i_d + i_{kd})$. An alternative way of instrumenting this nonlinearity on the computer, although the effect is identical, is to write equation 11 as

$$\omega_0\psi_{md} = X_{md}(i_f - i_d + i_{kd} - i_s) \tag{11B}$$

where i_s as a function of $\omega_0\psi_{md}$ and is obtained from the no-load magnetization curve as shown in Fig. 2. The X_{md} in equation 11(B) is the linear X_{md}. A complete computer diagram including a representation of saturation is given in Fig. 3. It is possible to reduce further the number of amplifiers required if any of the intermediate variables such as i_{kd} and i_{kq} are not desired. Care must be exercised in reducing the number of amplifiers not to violate the nonlinear equation 11(A) or 11(B).

The Phase Quantities

With the original stipulation that no zero-sequence current is required, the phase quantities are simply

$$v_a = v_d \cos\theta - v_q \sin\theta \tag{17}$$

$$i_a = i_d \cos\theta - i_q \sin\theta \tag{18}$$

where θ is the angle of the direct axis from the axis of phase a, measured in the direction of rotor rotation.

Therefore

$$v = \sqrt{v_d^2 + v_q^2} \tag{19}$$

and

$$i = \sqrt{i_d^2 + i_q^2} \tag{20}$$

where in the steady state v and i are equal to the rms value of the phase voltages and currents respectively. During a transient the phase quantities may contain d-c transients and thus the d and q components will contain variations at the alternator frequency. In order to obtain signals v and i which are equal to the peak or rms of a-c phase quantities during the transient as well as during the steady state, the ω_0 component should be removed from the d and q components before taking the square root of the sum of the squares. Thus depending upon the type of sensing in the regulator (high phase, average, etc.) the v_d and v_q signals may be used to generate a good approximation of the feedback signal to the regulator from the alternator. For paralleling and applying loads, the actual d and q signals are used, thus no approximation is necessary.

The equations for the load relations are obtained directly from equations 17 and 18. For example, for a balanced resistance-inductance (R-L) series load

$$v_a = (R + Lp)i_a \tag{21}$$

and thus

$$v_d = i_dR + X\frac{p}{\omega_0}i_d - X\frac{\omega}{\omega_0}i_q \tag{22}$$

$$v_q = i_qR + X\frac{p}{\omega_0}i_q + X\frac{\omega}{\omega_0}i_d \tag{23}$$

Representation of the Alternator Operating at No Load

From the point of view of representation on an analog computer, the external representation at the alternator load terminals for no load presented the greatest difficulty. Referring to Fig. 3 it is seen that the external connection must provide for the alternator the signals v_d and v_q given the signals i_d and i_q. This flow of signals is illustrated in Fig. 4. In the case of no load, this external connection must represent an infinite impedance, and thus in the computer this corresponds to an infinite gain. When a large gain was used at this point in an actual computer application, the noise level became excessively high for the computer and thus operation was not considered satisfactory.

There are several possible compromises

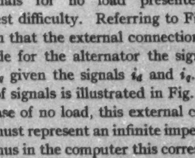

vd
vq
id
iq
EXTERNAL CONNECTION TO ALTERNATOR TERMINALS

Fig. 4. External connection to the alternator terminals

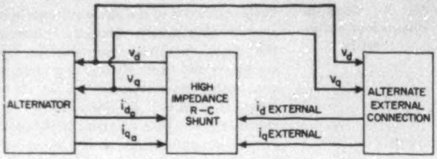


Fig. 5. Alternate external connection to the alternator terminals

in the representation to avoid this noise problem. First, the time base for the computer may be increased, thus reducing the effective gains in the computer loops and reducing the noise. Second, the no-load condition of the alternator may be approximated by a high-resistance load, since for a resistance load

$$v_d = i_d R \tag{24}$$

$$v_q = i_q R \tag{25}$$

Then, for example, if a $R=25$ per unit is used to approximate no load, the computer high gain is reduced and the noise level is correspondingly reduced. Thus, by using a large computer time base and approximating the open circuit with a reasonably high resistance, the computer noise is reduced to an acceptable level.

A further improvement may be obtained by approximating the no load condition by a high-impedance balanced shunt resistance-capacitance (R-C) load. The C may be considered to represent, at least in part, the distributed capacitances of the alternator. For this case

$$v_d = \frac{\omega_0}{p}\left(X_e i_d + v_q - v_d \frac{X_e}{R}\right) \tag{26}$$

$$v_q = \frac{\omega_0}{p}\left(X_e i_q - v_d - v_q \frac{X_e}{R}\right) \tag{27}$$

where

X_c = per-unit shunt capacitive reactance
R = per-unit shunt resistance
ω/ω_0 was approximated by 1

In this case, the external connection, although it has the properties of a high gain, has in addition the properties of a filter to the high-frequency noise of the computer. Thus, a lower time base or a higher impedance may be obtained to represent the no-load without having excessive noise. In the system involved in the computer study upon which this paper is based, a time base of 500 and for no load, $R=100$ per unit, $X_e=200$ per unit, were used without obtaining excessive noise level. It was found for the particular alternator being studied that $X_e=2R$ provided the best solution to the high-gain load problem, while for $X_e=R=100$ per unit a computer oscillation occurred.

If this high-impedance R-C shunt combination is considered a permanent part of the alternator, it is then possible to make external connections to the alternator with input signals of current and output signals of voltage as shown in Fig. 5. In many cases this is particularly convenient. In this case, the i_d and i_q of equations 26 and 27 are

$$i_d = i_{d\,\text{alt}} - i_d \text{ external connection} \tag{28}$$

$$i_q = i_{q\,\text{alt}} - i_q \text{ external connection} \tag{29}$$

For example, the case of an R-L series load which is to be suddenly applied or removed from the alternator is examined. The relations for this type of external connection are given by equations 22 and 23, rewritten here in a form to avoid differentiation for application to a computer.

$$i_d = \frac{\omega_0}{p}\left[\frac{v_d}{X} - i_d \frac{R}{X} + \left(\frac{\omega}{\omega_0}\right) i_q\right] \tag{22A}$$

$$i_q = \frac{\omega_0}{p}\left[\frac{v_q}{X} - i_q \frac{R}{X} - \left(\frac{\omega}{\omega_0}\right) i_d\right] \tag{23A}$$

The required input signals for such an external connection are v_d and v_q and the generated signals are i_d and i_q. The arrangement of Fig. 5 is therefore well adapted to this case, and closing the switch simulates application of the balance R-L load.

Parallel Operation

Up to this point, the representation of the alternator and external balanced loads have been treated. It is, of course, possible to set up on a computer more than one alternator with loads but the interconnection to represent paralleling of the alternator requires additional computation. The d and q components of the alternator are referred to its rotor angular position; and therefore, following the pattern of this paper, the two alternator representations may not be simply interconnected to represent paralleling since the d and q components of each are not referred to the same rotor angular position. If the equations of a second alternator are written with the d and q components referred to the rotor of a first alternator, these equations become very involved and much physical significance is lost.

A much better solution is obtained by simply allowing both alternator sets of

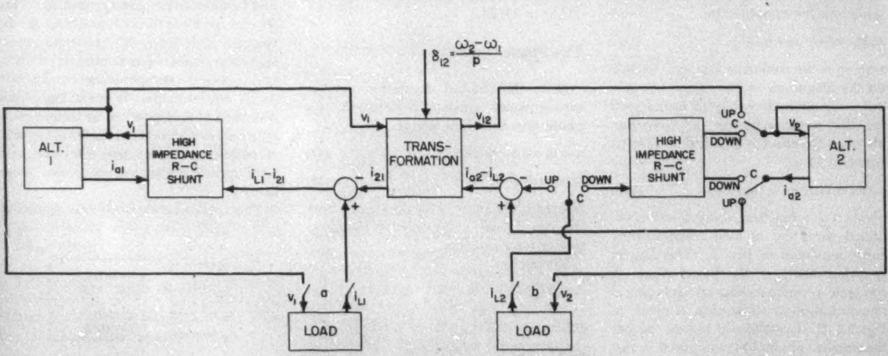


Fig. 6. Arrangement for paralleling two alternators

equations to be referred to the angular position of the rotor of each alternator and then transforming the external signals v_d, v_q, and i_d, i_q for interconnection to represent paralleling.

Thus, for the case of two alternators the angle δ_{12} between the two rotors is

$$\delta_{12}=\frac{1}{p}(\omega_2-\omega_1)$$

and the external signals are then transformed by

$$v_{d_{12}}=v_{d_1}\cos\delta_{12}+v_{q_1}\sin\delta_{12}$$

$$v_{q_{12}}=-v_{d_1}\sin\delta_{12}+v_{q_1}\cos\delta_{12}$$

$$i_{d_{21}}=i_{d_2}\cos\delta_{12}-i_{q_2}\sin\delta_{12}$$

$$i_{q_{21}}=i_{d_2}\sin\delta_{12}+i_{q_2}\cos\delta_{12}$$

where the subscript indicates the transformation. For example, $v_{d_{12}}$ indicates the transformation of the v_d from being referred to the rotor of alternator *1* to being referred to the rotor of alternator *2*. The resolver of an analog computer is well adapted to this transformation. A complete diagram indicating loading and transformation for paralleling two alternators is shown in Fig. 6. For diagram simplicity, only a single signal is shown for both the *d* and *q* components. Three switches are shown in the diagram. With switch *c* down, the two alternators are isolated and i_{21} will be equal to zero. Load is applied or removed from alternator *1* with switch *a* and alternator *2* with switch *b*. By switching *c* to the up position, a paralleling transient is initiated. With the switch *c* in the up position, the alternators are in parallel and load may be applied or removed from the parallel combination by using either switch *a* or *b*.

This procedure for paralleling is not limited to two machines, but any number may be paralleled. In general, one of the machines should be taken as the reference and only that machine will obtain the high-impedance R-C shunt loading after the paralleling. This reference machine with the high-impedance loading then provides the common voltages for all the paralleled machines through transformations such as $v_{d_{12}}$ and $v_{d_{13}}$. The current signals of the other alternators are then transformed such as $i_{d_{21}}$, $i_{d_{31}}$ to the reference machine where a summation of currents is used to determine the current for the reference machine.

Conclusions

In the computer study on which this paper is based, approximately 80 amplifiers plus resolvers, multipliers, and nonlinear function generators were required to represent two alternators, the associated constant-speed drives, excitation systems, and the interconnection for regulation of voltage magnitude and frequency and to obtain active and reactive load division. Tests were also made on the physical equipment. Test results are not given here since these results depend on the representation of the complete system of which the paralleled alternators are only a part. It is expected that these results will be treated in a future paper giving also the representation of the other system components.

In this paper the representation for parallel alternators has been treated. In the operation of an analog computer for this type of study, all transients are initiated by having the computer on "operate" and in the steady state and then closing or opening a switch. In switching from "reset" to "operate" some rather violet transients may ensue, but these are of no interest and the system must have settled into the steady state before any switches are operated to initiate the transients to be studied.

The same type of handling is good for any system of any number of parallel alternators providing there are no neutral connections. Even though this work is limited to the balanced passive load case, it is a straightforward extension to apply the same representation to any load, like an induction motor, which has a representation of the same type.

The two important extensions of this work are the case of any unbalanced load (with neutral connection) and a better representation for saturation. These extensions are presently under study.

References

1. SYNCHRONOUS MACHINES (book), Charles Concordia. John Wiley & Sons, Inc., New York, N. Y., 1951.
2. ANALOGUE COMPUTER REPRESENTATIONS OF SYNCHRONOUS GENERATORS IN VOLTAGE-REGULATION STUDIES, M. Riaz. *AIEE Transactions*, vol. 75, pt. III, Dec. 1956, pp. 1178–84.

◆

Discussion

G. W. Bills (North American Aviation, Inc., Downey, Calif.): This paper, just as was that by Riaz (see reference 2 of the paper), is timely for systems engineers in the guided missile field. At North American Aviation we are attempting to analyze parallel operation of alternators by separate digital and analog approaches. The digital method does not have the noise problems and circuit difficulties associated with electronic analog computers, but programming is difficult and the cost per study may be higher digitally. Have there been any attempts made by the authors to obtain digital solutions?

In the paper it is stated that equations 1–5 are in a per-unit system, but are not ω and ω_0 expressed in actual radians per second?

In the per-unit system in which the mutual inductances are the same for both the stator and rotor, was the per-unit field voltage v_f considerably smaller than 1.0? In attempting to make an analog representation of a synchronous machine we found that a normal v_f was less than 0.003 per unit, this value being the unscaled value.

Do the authors know of any objections to using resolvers made up from diode function generators (for square rooting) and electronic multipliers rather than resorting to excessive time scaling assuming that the no-load case is not being considered?

We hope this paper will be followed with another giving actual comparisons between test and analog results.

We wonder if there is any Wright Air Development Center (WADC) report on this effort aside from their Technical Report *55-311*.

L. V. Boffi and V. B. Haas, Jr.: We appreciate the discussion by Mr. Bills. It is constructive and indicates that he is working hard at the same problem and is interested in our results.

In regard to the questions raised by Mr. Bills, we submit the following:

1. We have not tried to solve this problem by use of a digital computer. Work is being done by the Energy Conversion Group at the Massachusetts Institute of Technology using digital computations, but the objective here is to optimize design rather than to obtain transient solutions.
2. Yes, ω and ω_0 are expressed in radians per second. The per-unit quantities are the ones enclosed in parenthesis like ω/ω_0 = per-unit speed and $\omega\psi$ = per-unit linkage, etc.
3. Yes, the per-unit field voltage v_f is much smaller than 1.0. It will be the same order of magnitude as r_f in per unit. Because of this fact it was found advantageous to use a capacitor larger than normal in the integrator following v_f.
4. In this study diode function generators and electronic multipliers were used along with servomultipliers and resolvers. We can see no objection to this. It is noted, however, that even when the no-load case is not being considered, one of the alternators (see Fig. 6) still requires a high-gain circuit. These high-gain circuits were the primary cause of the noise problem.
5. It is expected that this paper will be followed with another giving actual comparisons between test and analog results. Of necessity such a paper would include or must have reference to descriptions of the representations of all components in the

system such as the drives, regulators, and interconnections for active and reactive load division. Professor P. E. Smith, Jr. is presently preparing such a paper for the drives.

6. Several WADC reports contain information on this effort.[1–4]

It is expected that within a year a further report on the correlation between hardware tests and computer results for a complete system will be published

References

1. *WADC Technical Report 56-471*. Wright Air Development Center, Dayton, Ohio, Sept. 1956.

2. Study of Aircraft Alternating-Current Electric Power Generating Systems, Vinton B. Haas, Jr., Paul E. Smith, Jr., David C. White. *Ibid.*

3. *WADC Technical Note 56-384*. *Ibid.*, Sept. 1956.

4. Analog Computer Representations of Aircraft Parallel A-C Generating Systems, L. Boffi, V. B. Haas, Jr., M. Riaz, P. E. Smith, Jr. *Ibid.*

The Development of a New Method of Circuit Analysis in Ladder Networks

LAWRENCE F. COKER
STUDENT MEMBER AIEE

IT IS the purpose of this paper to present a method of circuit analysis which is simpler in form and can be more readily adapted to calculation than that of the conventional loop-voltage and nodal-current analyses. The method presented here is developed specifically for application to networks composed of L-sections in tandem. It is desired that relations between any two voltages in the circuit be obtained from inspection of the circuit; also, that similar relations be obtained for any two loop currents. Since the approach to the development of the method is purely empirical, the conclusions are drawn inductively. The results are derived directly through the use of Maxwell and Kirchoff rules of circuit analysis.[1,2]

Survey of the Problem

The difficulties encountered in numerical calculations through the use of nodal-current or loop-voltage analysis may be readily seen by considering Fig. 1, a tandem connection of three L-sections. It is desired to obtain an expression relating the input and output voltages, E_1 and V_6 respectively. Inspection reveals that three nodal voltages, V_2, V_4, and V_6, are unknowns requiring three equations and the solution of a third-order determinant. Similarly, there are three loop currents present, I_1, I_4, and I_3, the solution for which involves three mesh equations and a third-order determinant. It becomes apparent that each successive L-section connected to the output of the network introduces an additional nodal voltage and loop current, and the resulting solutions become increasingly tedious. Three nodal equations may be written:

$$Y_1E_1=(Y_1+Y_2+Y_3)V_2-Y_3V_4-0Y_6 \qquad (1)$$

$$0=-Y_3V_2+(Y_3+Y_4+Y_5)V_4-Y_5V_6 \qquad (2)$$

$$0=-0-Y_5V_4+(Y_5+Y_6)V_6 \qquad (3)$$

From which the determinant solution for V_6 becomes

$$V_6=\frac{\begin{matrix}(Y_1+Y_2+Y_3), & -Y_3 & , E_1Y_1\\ -Y_3 & ,(Y_3+Y_4+Y_5), & 0\\ 0 & , -Y_5 & , 0\end{matrix}}{\begin{matrix}(Y_1+Y_2+Y_3), & -Y_3 & , 0\\ -Y_3 & ,(Y_3+Y_4+Y_5), & -Y_5\\ 0 & , -Y_5 & ,(Y_5+Y_6)\end{matrix}} \qquad (4)$$

The solution for V_6 becomes:

$$V_6=\frac{Y_3Y_5E_1}{(Y_1+Y_2+Y_3)(Y_3+Y_4+Y_5)(Y_5+Y_6)-Y_3(Y_5+Y_6)-Y_5(Y_1+Y_2+Y_3)} \qquad (5)$$

The relation of input to output voltage becomes, after simplification,

$$\frac{E_1}{V_6}=\overset{(1)}{1}+\overset{(2)}{Z_1Y_2}+\overset{(3)}{Z_1Y_4}+\overset{(4)}{Z_3Y_4}+\overset{(5)}{Z_1Y_6}+\overset{(6)}{Z_3Y_6}+\overset{(7)}{Z_5Y_6}+\overset{(8)}{Z_1Z_3Y_2Y_4}+\overset{(9)}{Z_1Z_3Y_2Y_6}+\overset{(10)}{Z_1Z_5Y_2Y_6}+\overset{(11)}{Z_1Z_5Y_4Y_6}+\overset{(12)}{Z_3Z_5Y_4Y_6}+\overset{(13)}{Z_1Z_3Z_5Y_2Y_4Y_6} \qquad (6)$$

Survey of Results

Here it is seen that all impedances present in the expression are in series between the input and output and all admittances are in shunt between the input and output. This fact suggests a "summation of products of ratios of series to parallel impedances."

Referring to the circuit of Fig. 1, it is seen that E_1/V_6 will be given by the first two terms of equation 6 if each shunt impedance is opened with the exception of Z_2. Also, it is seen that E_1/V_6 will be given by the first term, unity, when all shunt impedances are opened.

These observations suggest a method of approach in determining the input to output relation for the entire circuit. First, all shunt admittances are opened and unit ratio between input and output is obtained:

$$\frac{E_1}{V_6}=1 \text{ (all } Y\text{'s open)} \qquad (7)$$

Each successive term in equation 6 contains shunt Y's and series Z's. Therefore, for greater clarity, all terms in equation 6 containing individual shunt Y's are collected. This results in:

$$\frac{E_1}{V_6}=1+Z_1Y_2+(Z_1+Z_3)Y_4+(Z_1+Z_3+Z_5)Y_6+Z_1Z_3Y_2Y_4+Z_1Z_3Y_2Y_6+Z_1Z_5Y_2Y_6+Z_1Z_5Y_4Y_6+Z_3Z_5Y_4Y_6+Z_1Z_3Z_5Y_2Y_4Y_6 \qquad (8)$$

Each of the terms in equation 8 containing individual Y's is now seen to be the product of all series impedances and the shunt admittance remaining when all shunt admittances, except the one involved in the term, are opened.

Since unity was obtained by opening all shunt Y's, and the next six terms in equation 6 were obtained by opening all shunt Y's but one, in turn, it follows that the next step should consist of opening all shunt Y's except two, in turn. Furthermore, it is suggested by the preceding results that the products of series impedances and shunt Y's should appear.

Since the eighth term in equation 6 involves only Y_2 and Y_4 it is reasonable to assume that this is the expression that results when Y_6 is opened. Fig. 2 shows the resulting circuit. The eighth term is then seen to be the product of two products

$$Z_1Z_3Y_2Y_4=(Z_1Y_2)(Z_3Y_4) \qquad (9)$$

It becomes apparent that the series im-

Paper 57-104, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted July 30, 1956; made available for printing January 3, 1957.

Lawrence F. Coker attends Texas Technological College, Lubbock, Tex.

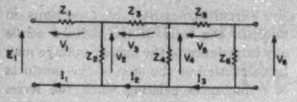


Fig. 1. Tandem connection of three L-sections

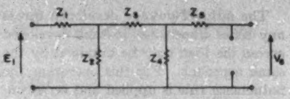


Fig. 2. Circuit of Fig. 1 with Y_6 opened

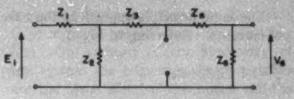


Fig. 3. Circuit of Fig. 1 with Y_4 opened

pedance associated with each shunt admittance is now restricted to the impedance existing between the shunt Y with which it is associated and the next adjacent Y nearest the source. A check reveals that all previously obtained ZY products are satisfied by this restriction.

The ninth and tenth terms involve Y_2 and Y_6 only, indicating that Y_4 is opened. The resulting circuit is shown in Fig. 3. Thus

$$Z_1Z_3Y_2Y_6+Z_1Z_5Y_2Y_6=Z_1Y_2(Z_3+Z_5)Y_6 \quad (10)$$

And by similar procedures, with Y_2 open, the 11th and 12th terms in equation 6 become

$$Z_1Z_3Z_5Y_2Y_4Y_6=Z_1Y_2Z_3Y_4Z_5Y_6 \quad (11)$$

Collecting equations 8, 9, 10, and 11, and combining them into a single equation results in

$$\begin{aligned}\frac{E_1}{V_6}=[1]+[&Z_1Y_2+(Z_1+Z_3)Y_4+\\&(Z_1+Z_3+Z_5)Y_6]+\\&[Z_1Y_2Z_3Y_4+Z_1Y_2(Z_1+Z_5)Y_6+\\&(Z_1+Z_3)Y_4Z_5Y_6]+[Z_1Y_2Z_3Y_4Z_5Y_6]\end{aligned} \quad (12)$$

Statement of the Voltage Method

Equation 12 is the final expression for the input to output voltage relationship, and may be written directly from the circuit without resort to voltage or current equations, or determinants. It offers the advantage of dimensionless computation and is especially convenient where series and shunt impedances are of the same order of magnitude. Another important feature is that it permits immediate knowledge of the frequency and phase characteristics of the circuit where resistance and capacitance only, resistance and inductance only, and pure reactance L-sections are placed in tandem.

For convenience, a new term is defined. A "reductance" is "the product of any shunt admittance occurring between the source and the load, and the series impedance occurring between that admittance and the next adjacent admittance nearest the source." As an example, refer to Fig. 4.

The reductance of Y_1 is zero since zero impedance exists between Y_1 and the source. The reductance of Y_3 is Z_2Y_3.

Thus, by the use of this definition, the previously outlined method may be stated as follows: The ratio of input to output voltage of any number of L-sections placed in tandem may be obtained by summing (1) the unity ratio obtained by opening all shunt admittances; (2) the sum of reductances obtained by opening all shunt admittances but one, in turn, until all have been included; (3) the sum of the products of reductances which are obtained by opening all but two shunt admittances, in turn, until all combinations are exhausted; (4) the sum of the products of reductances which are obtained by opening all but three shunt admittances, in turn, until all combinations, are exhausted; and (5), repeating step 4 for four admittances, and so on.

The Solution for Currents

An even simpler expression exists for the input to output current relationship. In this case, the method outlined previously leads immediately to the solution. Referring to Fig. 1, the loop voltage equations are obtained as

$$E_1=(Z_1+Z_2)I_1-Z_2I_2-0I_3 \quad (13)$$

$$0=-Z_2I_1+(Z_2+Z_3+Z_4)I_2-Z_4I_3 \quad (14)$$

$$0=0-Z_4I_2+(Z_4+Z_5+Z_6)I_3 \quad (15)$$

The determinant solution for the ratio of input to output current is

$$\frac{I_1}{I_3}=\frac{\begin{array}{lll}E & ,\ -Z_2 & ,\ 0\\ 0 & ,\ (Z_2+Z_3+Z_4), & -Z_4\\ 0 & ,\ -Z_4 & ,\ (Z_4+Z_5+Z_6)\end{array}}{\begin{array}{lll}(Z_1+Z_2), & -Z_2 & ,\ E_1\\ -Z_2 & ,\ (Z_2+Z_3+Z_4), & 0\\ 0 & ,\ -Z_4 & ,\ 0\end{array}} \quad (16)$$

The solution, when expanded, becomes

$$\frac{I_1}{I_3}=\frac{E_1(Z_2+Z_3+Z_4)(Z_4+Z_5+Z_6)-Z_4^2E_1}{Z_2Z_4E_1}$$

$$\frac{I_1}{I_3}=1+Y_2(Z_3+Z_5+Z_6)+Y_4(Z_5+Z_6)+Y_2Z_3Y_4(Z_5+Z_6) \quad (17)$$

Statement of the Current Method

The first term in equation 17 may be considered to apply to the ratio I_1/I_3 when only a single loop exists or when Y_2 and Y_4 are open. The next term applies when only Y_2 is present, and the third term when only Y_4 is present. The last term appears as a result of both Y_2 and Y_4 being present.

Thus another term, adductance, is defined. An "adductance" is "the product of the admittance of a shunt branch occurring between the input and the output current branches, and the series impedance between that branch and the next shunt branch nearest the output current branch."

In terms of the foregoing definition, the procedure may be stated as follows: The ratio of input to output current of any number of L-sections placed in tandem may be obtained by summing (1) unity; (2) the sum of adductances with shunt admittances taken one at a time, in turn; (3) the sum of the products of adductances with all shunt admittances taken two at a time, in turn; and (4) repetition of step 3 for shunt admittances taken four at a time, and so on. The series impedance of the output current branch is not considered a shunt branch since the output voltage is absent from the expression.

Other Applications and Conclusions

An interesting result occurs when a relationship between the output voltage and any nodal voltage, other than the source, is desired. Referring to Fig. 1, the following obtains:

$$\frac{V_2}{V_6}=1+Z_3Y_4+(Z_3+Z_5)Y_6+Z_3Y_4Z_5Y_6 \quad (18)$$

Also, following the reductance rule:

$$\frac{V_4}{V_6}=1+Z_5Y_6 \quad (19)$$

Thus any nodal voltage, between the source and load voltage, may be considered a source for that load voltage.

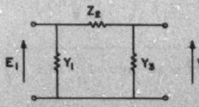


Fig. 4. Example of "reductance"

The same principle applied in the case of currents. Referring to Fig. 1,

$$\frac{I_3}{I_5}=1+Y_4(Z_5+Z_6) \qquad (20)$$

where in this case, I_3 is considered the source current, when the relation is obtained by the adductance rule.

Duality Features of the Method

Identical results are obtained by short-circuiting series impedances rather than opening shunt admittances. The definitions of reductance and adductance remain unaltered. As an example, Fig. 1 is again considered. The ratio E_1/V_6 is

$$\begin{array}{ll}
\text{Step 1:} & \dfrac{E_1}{V_6}=1 \text{ (short-circuiting all series impedances)}+\\
\text{Step 2:} & \left\{\begin{array}{l} Z_1(Y_2+Y_4+Y_6) \text{ (short-circuiting } Z_3 \text{ and } Z_5)+\\ Z_3(Y_4+Y_6) \text{ (short-circuiting } Z_1 \text{ and } Z_5)+\\ Z_5(Y_6) \text{ (short-circuiting } Z_1 \text{ and } Z_3)+ \end{array}\right.\\
\text{Step 3:} & \left\{\begin{array}{l} Z_1Y_2Z_3Y_4 \text{ (short-circuiting } Z_5)+\\ Z_1(Y_2+Y_4)Z_5Y_6 \text{ (short-circuiting } Z_3)+\\ Z_3Y_4Z_5Y_6 \text{ (short-circuiting } Z_1)+ \end{array}\right.\\
\text{Step 4:} & Z_1Y_2Z_3Y_4Z_5Y_6 \text{ (short-circuiting none)}
\end{array} \qquad (21)$$

The ratio of voltage developed across any series impedance to that developed across the load may be obtained by the same approach. For this operation, the reductance rule is applied and series impedances are short-circuited as described, with the exception of the impedance developing the series voltage. As an example, the ratio V_1/V_6 for Fig. 1 becomes

$$\begin{aligned}\frac{V_1}{V_6}=&Z_1(Y_2+Y_4+Y_6) \text{ (with } Z_3 \text{ and } Z_5 \text{ short-circuited)}+\\ &Z_1Y_2Z_3(Y_4+Y_6) \text{ (with } Z_5 \text{ short-circuited)}+\\ &Z_1(Y_2+Y_4)Z_5Y_6, \text{ (with } Z_3 \text{ short-circuited)}+\\ &Z_1Y_2Z_3Y_4Z_5Y_6 \text{ (with non short-circuited)}\end{aligned} \qquad (22)$$

Again the same principle applies in the case of currents, and the "short-circuiting" approach yields the same results as does the "opening" approach.

The form of the expression for the "short-circuiting" approach becomes

$$\begin{aligned}\frac{I_1}{I_5}=&1+Y_2Z_3+(Y_2+Y_4)(Z_5+Z_6)+\\ &Y_2Z_3Y_4(Z_5+Z_6)\end{aligned} \qquad (23)$$

Summary

In view of the preceding discussion, it is concluded that for any bilateral, passive circuit composed of L-sections in tandem, the ratio of the voltage across any impedance to the output voltage may be calculated as the sum of the products of reductances existing when the series impedances are considered singly, doubly, triply, and so on, with all possible combinations included. It is seen that identical results are obtained, except in the case of series voltage to output voltage ratios, by the sum of the products of reductances existing when shunt admittances are considered singly, doubly, and so on. The ratios of branch currents to output current are obtained by the same rules through the use of adductances rather than reductances.

Rigid mathematical proof of the generality of the foregoing analysis has not been obtained. By further study perhaps this method, with some modification, could be extended so as to apply to more complex circuits.

References

1. Résumé of Maxwell's and Kirchoff's Rules for Network Analysis, Y. H. Ku. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 253, no. 1, Jan. 1952, pp. 211–24.

2. Feedback Theory—Further Properties of Signal Flow Graphs, Samuel J. Mason. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 44, no. 7, July 1956, p. 920.

◆

Discussion

Pierre M. Honnell and **Donald F. Wann** (Washington University, St. Louis, Mo.): The author is to be complimented upon his intriguing method of circuit analysis for ladder-constructed networks, again illustrating the rich variety of special methods available for simple networks. As he suggests, a general method can be developed for more complex circuits.

If the elements of the impedance matrix $[Z]$ or the admittance matrix $[Y]$ of a network are known, then the current ratio or voltage ratio between two co-ordinates i and j with excitation on the k th may be found from

$$\frac{i_i}{i_j}=\frac{Z_{ki}}{Z_{kj}} \quad \text{and} \quad \frac{v_i}{v_j}=\frac{Y_{ki}}{Y_{kj}}$$

respectively, where the Z_{rs} and Y_{rs} are the cofactors of the respective impedance and admittance matrix determinants.[1] Furthermore, the elements of the $[Z]$ or $[Y]$ matrix may be determined from the network topology without the necessity of writing mesh or node equations, as Mason[2] illustrates. Since cofactors only are needed in this application, this yields the desired results quickly. Using this approach, the voltage or current ratios can be determined for dinode (i.e., linear, 2-node, lumped-element) constructed networks, and is not limited to networks having a common node.

It is clear, however, that dinode-constructed networks are a special case and general networks are normally encountered in practice. Thus, perhaps the matrix methods as originally introduced by Strecker and Feldtkeller[3] in 1929, using transmission matrices, present the better, more unified approach to this problem.

Using these techniques, linear networks including nondinode-constructed elements, such as vacuum tubes[4] or transistors may be systematically analyzed, yielding not only current and voltage, but also power. In addition, the recent publication of a generalized matrix stability criterion[5] presents the engineer with a powerful tool for analyzing networks containing closed feedback loops.

REFERENCES

1. Communication Networks, Vol. I (book), E. A. Guillemin. John Wiley & Sons, Inc., New York, N. Y., 1931, p. 151.

2. See reference 2 of the paper.

3. Theory of the Generalized Quadrupole, F. Strecker, R. Feldtkeller. *Elektrische Nachrichten-Technik*, Berlin, Germany, vol. 6, Mar. 1929, pp. 93–112.

4. The Application of Matrices to Vacuum-Tube Circuits, J. S. Brown, F. D. Bennett. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 37, July 1948, pp. 844–52.

5. The Generalized Transmission Matrix Stability Criterion, Pierre M. Honnell. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 292–98.

Transient Analysis of Second-Order Flip-Flops

L. M. VALLESE
ASSOCIATE MEMBER AIEE

Synopsis: A procedure for the analysis of the trigger sensitivity and of the switching speed of flip-flops of second order is developed. In particular the phase plane topology is investigated and the transition time evaluated. The method is applied first to a system with the simplest possible differential equation, then extended to a point contact emitter input transistor flip-flop.

THE characteristics of flip-flops of greatest interest to the designer are trigger sensitivity, switching speed, and reliability of operation. The first is expressed in terms of the minimum required amplitude of the trigger pulse, the second is related to the sum of the triggering and of the autonomous (undriven) transition times, the third is a function of the reliability of the various circuit components. It is the purpose of this paper to derive approximate analytical relationships between the flip-flop parameters and the flip-flop sensitivity and speed characteristics.

Several investigations of the dynamic behavior of flip-flops have appeared in the recent literature.[1-9] Most of these are concerned with the behavior in the mid-region of the transition interval, which is assumed to be linear; other investigations, concerned with the over-all transition, use the technique of phase paths which correlate together two voltages or a voltage and a current, etc. The technique of phase motions (correlating a variable and its time derivative), which will be used in this paper, does not seem to have been considered so far for this application. Although the two phase plane procedures are fundamentally equivalent, the latter one is more general, since it does not depend on the choice of the variable, and it is more useful, as it permits a direct computation of the transition time, when the equation of the trajectory may be expressed explicitly.

For convenience, the analysis is developed first for the simplest possible flip-flop system (heretofore called canonical flip-flop); the results are then generalized for application to practical cases. Eccles-Jordan flip-flops, which are systems of fourth order, will not be considered in this paper.

Analysis of Canonical Flip-Flop

A flip-flop may be defined in general as a nonlinear system possessing two stable singularities separated by a saddle point. Assuming that these are located at $(-x_1, 0)$, $(0,0)$, $(x_1, 0)$ in the phase plane, it is possible to synthesize the simplest differential equation possessing these prescribed singularities[10,11]; this is found to be

$$\ddot{x}+2b\dot{x}-ax(1-x^2/x_1^2)=0 \tag{1}$$

where a and b are positive quantities. According as

$$b \lesssim \sqrt{2a} \tag{2}$$

the stable singularities are respectively foci or nodes. In equation 1 the expression $f(x)=-ax(1-x^2/x_1^2)$ represents the so-called static characteristic of the flip-flop, with slopes $+2a$, $-a$, $+2a$ at the singular points. For purposes of simplification it is convenient to replace $f(x)$ in equation 1 with its piecewise linear approximation, Fig. 1, i.e., with

$$\begin{aligned} f_1(x)&=2a(x+x_1) \text{ for } x\leqslant -x_m\\ &=-ax \text{ for } -x_m\leqslant x\leqslant x_m\\ &=2a(x-x_1) \text{ for } x_m\leqslant x \end{aligned} \tag{3}$$

where $x_m=2x_1/3$. Correspondingly, the differential equation 1 is replaced with a piecewise linear differential equation of type

$$\ddot{x}+2b\dot{x}+f_1(x)=0 \tag{4}$$

The topological configuration of the phase plane of equation 4 (see Appendix I) is shown in Fig. 2 for the case $b<\sqrt{2a}$ and in Fig. 3 for the case $b>\sqrt{2a}$. In either case the line through the origin with slope $m=-(b+\sqrt{a+b^2})$ divides the phase plane in two regions in each of which all the trajectories terminate at the stable singularity enclosed. So that triggering may occur therefore, the point (x,y) must be moved across this dividing line. On account of the symmetry of the system the triggering characteristics are the same whether the initial state is $(-x_1,0)$ or $(x_1,0)$; in the following analysis triggering is considered from $(-x_1,0)$.

If triggering is obtained by application of a rectangular pulse of amplitude S and duration τ, the differential equation 4 is correspondingly modified for the duration τ assuming the form

$$\ddot{x}+2b\dot{x}+f_1(x)-S=0 \tag{5}$$

Comparison with equation 4 shows that the nature of the singularities does not change during triggering; but their location is moved from $(-x_1,0)$, $(0,0)$, $(x_1,0)$, to $(-x_1+S/2a,0)$ $(-S/a,0)$, $(x_1+S/2a,0)$ respectively. Figs. 4 and 5 are qualitative examples of the phase trajectories which are obtained in the case $b<\sqrt{2a}$ and $b>\sqrt{2a}$ respectively. It is of interest to note that the displacements of the singular points occur in such a direction as to co-operate in aiding the triggering operation.

For constant trigger amplitude the over-all transition time of the flip-flop increases with $b/\sqrt{2a}$; on the other hand, if $b/\sqrt{2a}<1$, switching is accompanied by overshoot. For these reasons it appears that the most satisfactory operation occurs when $b=\sqrt{2a}$. Accordingly, the triggering sensitivity and the triggering and over-all transition times will be computed first for this condition and then generalized to the case $b>\sqrt{2a}$.

From Fig. 5 it is seen that the minimum necessary amplitude of the trigger pulse is

$$S_{\min}=2ax_1/3 \tag{6}$$

which is the peak value of $f_1(x)$. On the other hand, it is of interest to point out that in the case $b<\sqrt{2a}$, Fig. 4, the minimum necessary amplitude is less than the value of equation (6).

For given amplitude $S\geqslant S_{\min}$ the minimum required duration of the trigger pulse must be such as to move the point (x,y) of Fig. 5 across the critical dividing line, i.e., from $x=x_1$, to $x=-x_s$. Assuming that $x_s=0$, one finds in nondimensional form (see Appendix II):

$$\tau_s=\sqrt{2a}=(\tau_{s_1}+\tau_{s_2})\sqrt{2a}\cong 2\left(1+\sqrt{\frac{3S}{ax_1}-1}\right)\Bigg/\left(\frac{3S}{ax_1}-2\right)+4.44 \ln\left(1+\frac{1}{3.34\dfrac{S}{ax_1}1-\exp(-\tau_{s_1}\sqrt{2a})-2.22}\right) \tag{7}$$

In equation 7 τ_{s_1} and τ_{s_2} are the fractional

Paper 57-81, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21-25, 1957. Manuscript submitted June 4, 1956; made available for printing December 4, 1956.

L. M. VALLESE is with Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

The work described was supported by the Office of Naval Research under contract with the Microwave Research Institute of the Polytechnic Institute of Brooklyn.

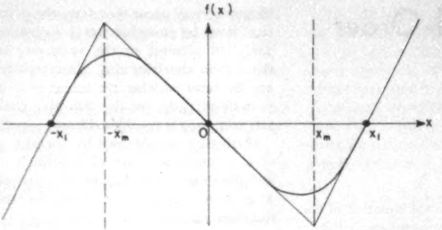


Fig. 1. Static characteristic of canonical flip-flop

transition times from $x=-x_1$ to $x=-x_m$ and from $x=-x_m$ to $x=-x_s$ respectively. These quantities have been plotted in Fig. 6 versus the normalized trigger pulse amplitude. It is of interest to note that τ_{s_1} is larger than τ_{s_2}, that τ_s is infinite when $S/ax_1=2/3$ (condition of minimum triggering amplitude) and decays rapidly as S/ax_1 increases.

Assuming that the application of the trigger pulse lasts only for the minimum time given by equation 7, the further motion of the flip-flop occurs autonomously. The corresponding transition time from $x=-x_s$, to $x=x_1$ is found to be approximately (see Appendix II) in nondimensional form as given in the following equation

$$\tau_a\sqrt{2a}=(\tau_{a_1}+\tau_{a_2})\sqrt{2a}\cong 4.45\ln\frac{8.85}{2.71\dfrac{S}{ax_1}(1-\bar{e}^{\tau_{s_1}\sqrt{2a}})-1}+2.302 \tag{8}$$

In equation 8 τ_{a_1} and τ_{a_2} are respectively the fractional transition times from $x=-x_s$ to $x=x_m$ and from $x=x_m$ to $x=x_1$. In Fig. 7 τ_{a_1} and τ_{a_2} have been plotted versus S/ax_1 in nondimensional form. It is noted that τ_{a_1} is larger than τ_{a_2}, that τ_{a2} is (approximately) independent of S/ax_1, and that τ_a decreases with the increase of S/ax_1. The reason for the approximately constant τ_{a_2} value is that the trajectories of the mid-region approach asymptotically the same straight line.

A comparison of the relative values of τ_s and τ_a is obtained in Fig. 8 where these quantities as well as the over-all transition time are plotted versus S/ax_1. It is interesting to note that τ_a is larger than τ_s.

The previous analysis has shown that it is possible to increase the switching speed by increasing a; however, if the triggering sensitivity is kept constant, the product ax_1 must be invariant and the separation of the two stable singular points is decreased correspondingly. A limitation in this increase of the speed is imposed by various reliability considerations.

When the coefficient b is larger than $\sqrt{2a}$ approximate expressions for the transition times may be derived similarly. In terms of the parameter $h=b/\sqrt{2a}$ the triggering time is approximately in nondimensional form given in the following equation

$$\tau_s\sqrt{2a}=(\tau_{s_1}+\tau_{s_2})\sqrt{2a}\cong\frac{1}{h-\sqrt{h^2-1}}\times \ln\left[\left(1+\frac{h}{\sqrt{h^2-1}}\right)\Big/2\left(1-\frac{2ax_1}{3S}\right)\right]+ \frac{1}{-h+\sqrt{h^2+1/2}}\ln\left[1-\frac{2x_1\sqrt{2a}}{3y(\tau_{s_1})}\times(-h+\sqrt{h^2+1/2})\right] \tag{9}$$

where

$$y(\tau_{s_1})\cong\frac{S}{\sqrt{2a}}\left[h\left(1-\frac{2ax_1}{3S}\right)-\frac{1}{2}(h+\sqrt{h^2-1})\times \exp(-h+\sqrt{h^2-1})\tau_{s_1}\sqrt{2a}\right]$$

Similarly, the autonomous transition time is in nondimensional form

$$\tau_a\sqrt{2a}=(\tau_{a_1}+\tau_{a_2})\sqrt{2a}\cong\frac{1}{-h+\sqrt{h^2+1/2}} \ln\frac{\dfrac{4x_1}{3}\sqrt{h^2+1/2}}{\dfrac{y(\tau_{s_1})}{\sqrt{2a}}+\dfrac{2x_1}{3}\left(-h+\sqrt{h^2+\dfrac{1}{2}}\right)}+ \frac{1}{h-\sqrt{h^2-1}}\ln 5\left[1+\frac{h}{\sqrt{h^2-1}}-3y(\tau_{a_1})/x_1\sqrt{2a}\sqrt{h^2-1}\right] \tag{10}$$

where

$$y(\tau_{a_1})\cong-\frac{2hx_1\sqrt{2a}}{3}+\left[\frac{y(\tau_{s_1})}{2}+\left(\sqrt{h^2+\frac{1}{2}}-h\right)\frac{x_1\sqrt{2a}}{3}\right]\times \exp\left[\left(\sqrt{h^2+\frac{1}{2}}-h\right)\tau_{a_1}\sqrt{2a}\right]$$

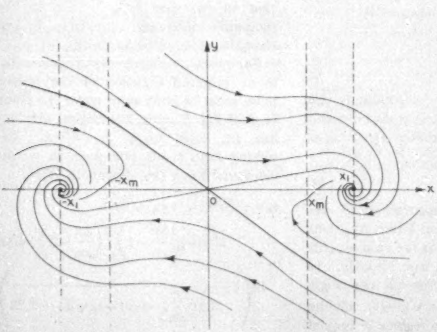


Fig. 2. Topological configuration of the phase plane for $b<\sqrt{2a}$

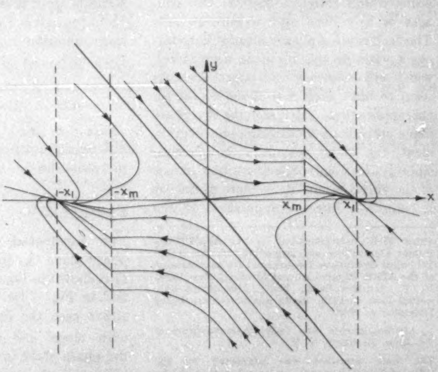


Fig. 3. Topological configuration of the phase plane for $b>\sqrt{2a}$

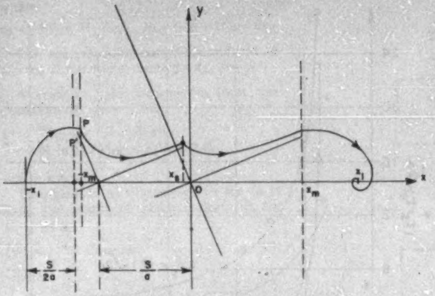


Fig. 4 (above). Example of phase trajectory for $b<\sqrt{2a}$

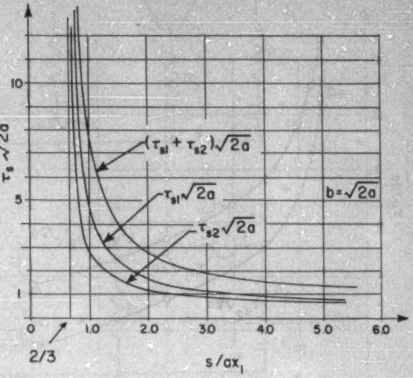


Fig. 6 (right). Triggering time versus trigger amplitude

Point-Contact Transistor Flip-Flop

A basic switching circuit is that obtainable with a single point-contact transistor, connected to present a nonlinear static input resistance characteristic with a negative intermediate range. The emitter input configuration will be investigated in the following as an example of application of the previous method of analysis. This circuit has been analyzed in references 3, 7, and 8.

The flip-flop is shown in Fig. 9 together with its approximate d-c equivalent circuit. This is drawn neglecting all intrinsic reactive elements but that of the internal voltage generator, and assuming that the resistances r_e, r_b, r_c, r_{mo} vary discontinuously from one constant value to another in going from the "cutoff" to the "active" and to the "on" regions; see Fig. 9(C). For the internal voltage generator the following differential equation is assumed

$$\tau\frac{dv}{dt}+v=r_{mo}i_e$$

with $\tau = 1/\omega_{co}$, or in operational form

$$v=\frac{r_{mo}}{1+\tau p}i_e \tag{11}$$

On the basis of the equivalent circuit the behavior of the flip-flop is found to be described by the following nonlinear differential equation (neglecting minority carrier storage effects)

$$\frac{d^2ie}{dt^2}+k_1\frac{di_e}{dt}+k_2i_e=k_3 \tag{12}$$

where the coefficients, all variable from one region to the other, assume the values given in the following. Let

$$R_p=R_b(R_c+r_c)/(R_b+R_c+r_c),\ R_i'=R_p+r_e$$

$$R_i''=R_p[1-r_{mo}/(R_c+r_p)],$$

$$R_i'''=R_bR_c/(R_b+R_c)$$

then equation 12 in the three linear regions is written, respectively as follows: In the cutoff region

$$p^2i_e+2b'\dot{i}_e+2a'i_e=-k_3' \tag{13}$$

where

$$2b'=\frac{1}{C_1}\left(\frac{1}{R_i'}+\frac{1}{R_e}\right)+\frac{1}{\tau}$$

$$2a'=[R_p+R_e+r_e]/R_eCR_i'\tau$$

$$k_3'=\left[\frac{-R_bE_{cc}}{R_b+R_c+r_c}+E_{ee}\right]/R_eCR_i'\tau$$

In the active region

$$p^2i_e-2b''pi_e-a''i_e=-k_3'' \tag{14}$$

where

$$2b''=\frac{|R_i''|}{R_p\tau}-\frac{1}{C_1}\left(\frac{1}{R_p}+\frac{1}{R_e}\right)$$

$$a''=[|R_i''|-R_e]/R_eCR_p\tau$$

$$k_3''=\{E_{ee}-R_bE_{cc}/[R_b+R_c+r_c]\}/R_eCR_p\tau$$

In the on region

$$p^2i_e+2b'''pi_e+2a'''i_e=k_3''' \tag{15}$$

where

$$2b'''=\frac{1}{C_1}\left(\frac{1}{R_i'''}+\frac{1}{R_e}\right)+\frac{1}{\tau}$$

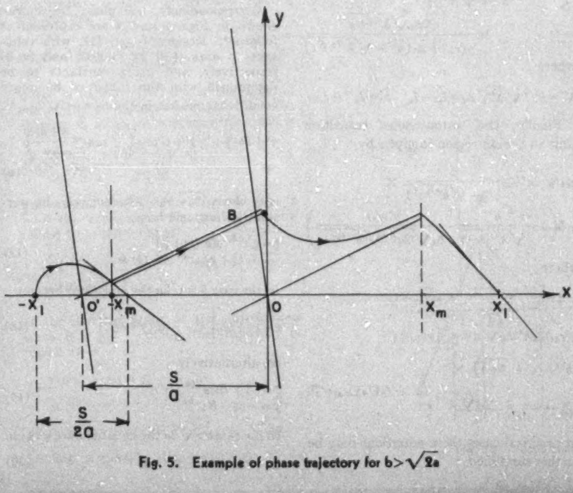


Fig. 5. Example of phase trajectory for $b>\sqrt{2a}$

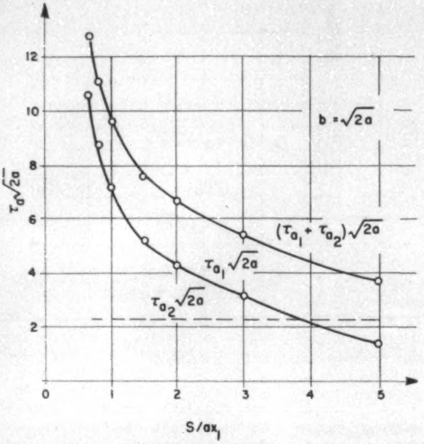


Fig. 7. Autonomous transition time versus trigger amplitude

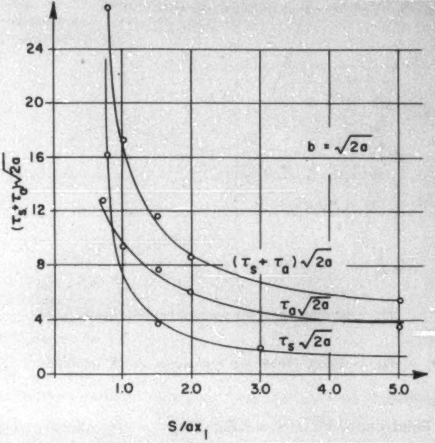


Fig. 8. Over-all transition time versus trigger amplitude

$$2a'''=[R_i'''+R_e]/R_eCR_i'''\tau$$

$$k_1'''=\left[\frac{E_{cc}}{R_c}-\frac{E_{ee}}{R_i'''}\right]/R_eC\tau$$

A comparison of the differential equations 4 and 12 is very instructive. The static characteristic of equation 12, shown in Fig. 10, has different slopes in the cutoff and in the on regions and therefore correspondingly different switching speeds. The singular points occur at

$$I_{e1}=\frac{-k_1'}{2a'},\ I_{e2}=\frac{k_1''}{a''},\ I_{e3}=\frac{k_1'''}{2a'''}$$

and the abscissas of separation of the linear regions at, respectively,

$$I_e'=\frac{k_1''-k_1'}{2a'+a''},\ I_e''=\frac{k_1''+k_1'''}{a''+2a'''}$$

The coefficient k_1 also varies nonlinearly; in the cutoff and in the on regions it has a value much larger than critical but, on the other hand, in the active region it becomes negative, thus providing, instead than subtracting, energy. The phase plane topology is not much modified, but the slopes of the asymptotes in the active region are now

$$m'=b''-\sqrt{a''+b''^2},\ n'=b''+\sqrt{a''+b''^2}$$

i.e., m' is a small negative number, and n' a large positive number. Correspondingly the triggering time τ_{s2} is generally negligible and the rate of change of i_e reaches very high values. The expressions of the transition times are obtained with minor modifications from equations 9 and 10. For example, for transition from cutoff, let $h'=b'/\sqrt{2a'}$ and $\Delta_1=I_e'-I_{e1}$, and the triggering time is given by

$$\tau_{s1}\sqrt{2a'}\cong\frac{1}{h'-\sqrt{h'^2-1}}\times \ln\left[\left(1+\frac{h'}{\sqrt{h'^2-1}}\right)\Big/2\left(1-\frac{2a'\Delta_1}{S}\right)\right]$$

Similarly the autonomous transition time in the active region is given by

$$\tau_{a1}\sqrt{a''}\cong\frac{1}{h''+\sqrt{h''^2+1}}\times \ln\frac{2\Delta_3\sqrt{h''^2+1}}{y(\tau_{s1})+\Delta_2\left(h''-\sqrt{h''^2+1}\right)}$$

where

$$h''=b''/\sqrt{a''},\ \Delta_2=I_{e2}-I_e',\ \Delta_3=I_e''-I_{e2}$$

Finally, the autonomous transition time in the on region is given by

$$\tau_{a2}\sqrt{2a'''}\cong\frac{1}{h'''-\sqrt{h'''^2-1}}\times \ln 5\left(1+\frac{h'''}{\sqrt{h'''^2-1}}-\frac{y(\tau_{a1})}{\Delta_4\sqrt{2a'''}\sqrt{h'''^2-1}}\right)$$

where

$$h'''=b'''/\sqrt{2a'''},\ \Delta_4=I_{e3}-I_e''$$

$$y(\tau_{a1})\cong h''\Delta_3\sqrt{a''}+\frac{1}{2}\left[y(\tau_{s1})+\left(h''-\sqrt{h''^2+1}\right)\times \Delta_2\sqrt{a''}\right]e^{\left(h''+\sqrt{h''^2+1}\right)\tau_{a1}\sqrt{a''}}$$

In practical cases these equations may be further simplified.

Appendix I. Phase Plane Configurations of Equation 4

The phase plane differential equations corresponding to equation 4 are

$$\frac{dy}{dx}=\frac{-2a(x+x_1)-2by}{y};\ \text{region I},\ x<-x_m$$

$$\frac{dy}{dx}=\frac{ax-2by}{y};\ \text{region II},\ -x_m\le x\le x_m$$

$$\frac{dy}{dx}=\frac{-2a(x-x_1)-2by}{y};\ \text{region III},\ x_m\le x$$

Correspondingly, the phase trajectories shown in Figs. 2 and 3 are expressed as follows[13]: Regions I and III, with reference to axes (ξ,y) at $(-x_1,0)$ and $(x_1,0)$ respectively, and using constants to be determined with the initial or boundary conditions, one has in the case $b<\sqrt{2a}$

$$y^2+2b\xi y+2a\xi^2=C\exp\left(\frac{2b}{k}\tan^{-1}\frac{y+b\xi}{ky}\right) \tag{16}$$

with $k=\sqrt{2a-b^2}$. Alternatively in parametric form one has

$$\begin{aligned}\xi&=\xi_0\epsilon^{-bt}\sin(kt+\varphi)\\ y&=-b\xi+k\xi_0\epsilon^{-bt}\cos(kt+\varphi)\end{aligned} \tag{17}$$

In the case $b=\sqrt{2a}$ the equations become

$$\xi=(y+b\xi)\left[C-\frac{1}{b}\ln(y+b\xi)\right] \tag{18}$$

or, alternatively

$$\begin{aligned}\xi&=(A+Bt)\epsilon^{-bt}\\ y&=-b\xi+B\epsilon^{-bt}\end{aligned} \tag{19}$$

In the case $b>\sqrt{2a}$ the equations are written

$$(y-n_1\xi)^{-n_1}=C(y-n_2\xi)^{-n_2} \tag{20}$$

where

$$n_1=-b+\sqrt{b^2-2a},\quad n_2=-(b+\sqrt{b^2-2a})$$

The corresponding curves are tangent to $y=n_1\xi$ at the origin and to $y=n_2\xi$ at $\xi=\infty$.

Alternatively in parametric form the equations 20 are written

$$\begin{aligned}\xi&=\epsilon^{-bt}(A\,Sh\,\eta t+B\cosh\eta t)\\ y&=-b\xi+\eta\epsilon^{-bt}(A\cosh\eta t+B\,Sh\,\eta t)\end{aligned}\qquad(21)$$

where $n=\sqrt{b^2-2a}$.

In region II with reference to the (x,y) axes, the family of phase trajectories for $b\gtrless\sqrt{2a}$ is

$$(y-nx)^{-n}=C(y-mx)^{-m}\qquad(22)$$

where

$$m=-(b+\sqrt{a+b^2}),\ n=-b+\sqrt{a+b^2}$$

The curves are asymptotic to $y=mx$ and to $y=nx$. In parametric form, equations 21 apply, provided $n=\sqrt{a+b^2}$.

Appendix II. Computation of Transition Times

In the case $b=\sqrt{2a}$ the trajectory in region I is

$$\xi=-\frac{S}{2a}(1+\sqrt{2a}\,t)\,\epsilon^{-\sqrt{2a}\,t}$$

$$y=-\sqrt{2a}\,\xi-\frac{S}{\sqrt{2a}}\,\epsilon^{-\sqrt{2a}\,t}$$

The triggering time τ_{s_1} from $\xi=-S/2a$ to $\xi=-S/2a+x_1/3$ follows from

$$1-\frac{2ax_1}{3S}=(1+\sqrt{2a}\,\tau_{s_1})\,\epsilon^{-\sqrt{2a}\,\tau_{s_1}}$$

An approximate solution is given in equation 7.

The trajectory in region II may be obtained from equation 21 or 22. As an approximation it may be replaced with a straight line

$$y=y(\tau_{s_1})+n\left(\xi-\frac{S}{a}+\frac{2x_1}{3}\right)$$

where $n=(\sqrt{3}-\sqrt{2})\sqrt{a}$ is the slope of one of the asymptotes of region II. The abscissa x_s at which triggering may end is

$$X_s\cong\frac{y(\tau_{s_1})}{2\sqrt{3}}+\frac{x_1(\sqrt{3}-\sqrt{2})}{2\sqrt{3}}$$

For example, if triggering lasts up to $x=0$ the expression of equation 7 is obtained from the integral

$$\tau_{s_2}=\int_{S/a-\frac{2x_1}{3}}^{S/a}\frac{d\xi}{y}$$

If the starting point of the autonomous state is $x=0$, $y=y(\tau_{s2})$, the trajectory in region II is

$$x=\frac{y(\tau_{s_2})}{\sqrt{3a}}\,\epsilon^{-\sqrt{2a}\,t}\,Sh\sqrt{3a}\,t$$

$$y=-\sqrt{2a}\,x+y(\tau_{s_2})\epsilon^{-\sqrt{2a}\,t}\cosh\sqrt{3a}\,t$$

Letting $x=2x_1/3$, the expression of τ_{a_1}

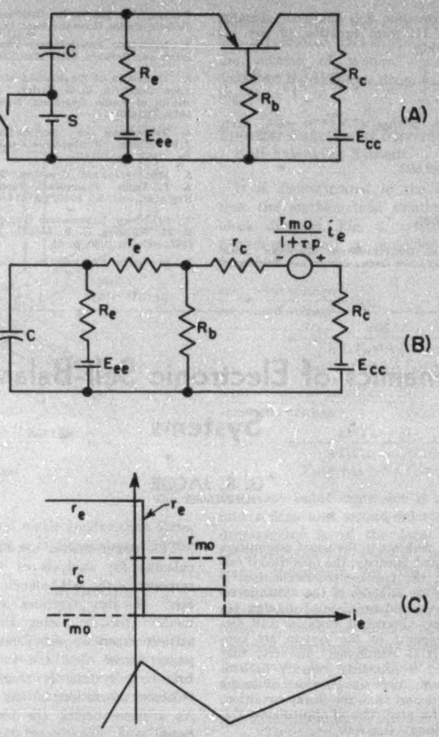


Fig. 9. Emitter input transistor flip-flop

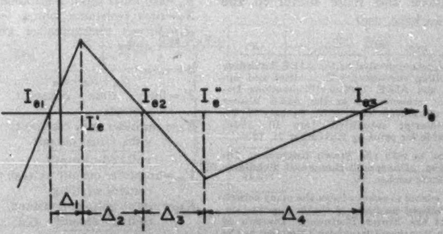


Fig. 10. Static characteristic of emitter input flip-flop

given in equation 8 is obtained. Finally, in region III from equation 19 one has the following

$$\xi=\left\{-\frac{x_1}{3}+\left[y(\tau_{a_1})-\sqrt{2a}\,\frac{x_1}{3}\right]t\right\}\epsilon^{-\sqrt{2a}\,t}$$

wherefrom, letting $\epsilon^{-\sqrt{2a}\,\tau_{a_2}}=0.1$, there follows

$$\sqrt{2a}\,\tau_{a_2}\cong 2.302$$

References

1. Uniform Relationships Between Sinusoids, Relaxation Oscillations and Triggered Pulses, W. Reichardt. *Elektrische Nachrichten-Technik*, Berlin, Germany, Sept. 1943, pp. 213–25.

2. Negative Resistances, Their Characteristics and Effects, *Ibid.*, Mar. 1943, pp. 76–87.

3. Dynamics of Transistor Negative Resistance Circuits, B. G. Farley. *Proceedings*, Institute of Radio Engineers, New York, N. Y., 1952, p. 1497.

4. Transition of Eccles-Jordan Circuit, J. R. Tillman. *Westinghouse Engineer*, Pittsburgh, Pa., 1951, pp. 101–10.

5. Multivibrator Circuits, M. V. Kiebert, A. F. Inglis. *Proceedings*, Institute of Radio Engineers, vol. 33, 1945, pp. 534–39.

6. Speed of Electronic Switching Circuits, E. M. Williams, D. F. Aldrich, J. B. Woodford. *Ibid.*, vol. 38, 1950, p. 65.

7. Stability Analysis of a Basic Transistor Switching Circuit, T. R. Bashkow. *Proceedings* National Electronics Conference, Chicago, Ill., 1954

8. The Transient Response to Transistor Switching Circuits, L. Lebow, R. H. Baker, R. E. McMahon. *Technical Report 27*, Massachusetts Institute of Technology, Cambridge, Mass., 1953.

9. Nonsaturating Pulse Circuits Using Two Junction Transistors, J. G. Linvill. *Proceedings*, Institute of Radio Engineers, vol. 43, 1955, p. 826.

10. On the Synthesis of Nonlinear Systems, L. M. Vallese. *Proceedings*, Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., 1953, p. 201.

11. A Note on the Analysis of Flip-Flops, L. M. Vallese. *Ibid.*, 1956.

12. Theory of Oscillations (book), A. A. Andronow, C. E. Chaikin. Princeton University Press, Princeton, N. J., 1949.

Dynamics of Electronic Self-Balancing Systems

G. R. JACOB
NONMEMBER AIEE

Synopsis: Subject to the usual assumption of component linearity, the operational calculus, i.e., the Laplace transformation, is applied to the analysis of the dynamic response of the industrial self-balancing potentiometer. Transient response and frequency response of the system are considered. It is established, however, that the system is primarily velocity limited. For large amplitude signals which cause the system to depart from the linear condition, a method for prediction of approximate frequency response is given.

IN THIS AGE of automation and intensified scientific research, evaluation of the dynamic response of the electronic self-balancing system is the key to good instrumentation. The development of the self-balancing system has preceded the wide adoption of the methods of operational calculus. However, with the increased need for better dynamics, such methods of evaluation are becoming more and more useful to the designer and the user.

This paper makes use of operational calculus for analysis of the dynamic response of the self-balancing potentiometer. It also describes an analytical method for expressing the effects of saturation on frequency response. The paper shows that the industrial self-balancing system is characterized by velocity saturation of the servomotor. As a consequence, the limits of nonlinear and linear operation must be determined; methods for this separation of linear from nonlinear response are given. Step inputs, ramp inputs, and sinusoidally varying inputs are considered.

Nomenclature

V = desired value, or command input from voltage source
R = reference input
C = controlled variable or slidewire voltage
B = feedback voltage
E = actuation error
R_a = amplifier input resistance, ohms
R_d = filter resistance, ohms
R_s = output resistance of voltage source, ohms
A = ratio, $\dfrac{R_a}{R_a+R_d+R_s}$
T_1 = filter time constant, $(R_d+R_s)C_d$; seconds
K_a = amplifier gain constant, $\dfrac{\text{volts (rms) output}}{\text{volts (d-c) input}}$
V_m = amplifier output voltage to servomotor control winding
K_m = motor voltage constant, $\dfrac{\text{radians/second}}{\text{volts (rms)}}$
T_m = motor and load time constant, seconds
K_{sw} = slidewire gradiant, volts/radian
$N=K_aK_mK_{sw}$
θ_m = angular position of motor pinion
ζ = conventional damping ratio associated with second-order transfer function
ω_1 = conventional frequency of oscillation associated with second-order transfer function
$V_{v\text{-sat}}$ = saturation velocity referred to the slidewire, volts/second
V_{sat} = minimum magnitude of step change in input level which results in $V_{v\text{-sat}}$, volts
V_R = velocity of ramp change in input level, volts/second
t_{sat} = time at which instantaneous velocity of sine-wave input is equal to instrument saturation velocity, seconds
α = angle corresponding to t_{sat}, radians
t_{lin} = time at which controlled variable returns to linear condition, seconds
ϕ = phase lag of controlled signal peak amplitude behind the command signal peak amplitude, for an input frequency at which system saturation just occurs, degrees
f_{sat} = frequency of sinusoidal change in input level at which system saturation just occurs

Physical Representation

For discussion purposes, the self-balancing system is defined as an application of the type-*1* servomechanism system[1] to a position-control device which results in a zero-displacement-error system under steady-state conditions. A familiar example of this system is the industrial recorder and voltage source as shown in Fig. 1.

In its simplest form, the system consists of a voltage source and a recorder. A slidewire and a battery, a chopper modulated amplifier, a 2-phase servomotor, and a filter and damping network comprise the recorder. The filter and damping network (shown by dashed line in Fig. 1) may be omitted where its function can be provided elsewhere (e.g., inherently in the amplifier design).

Paper 57-17, recommended by the AIEE Recording and Controlling Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N.Y., January 21–25, 1957. Manuscript submitted July 10, 1956; made available for printing November 21, 1956.

G. R. Jacob is with the Brown Instrument Division of the Minneapolis-Honeywell Regulator Company, Philadelphia, Pa.

The author wishes to acknowledge the many contributions of various members of the engineering department of the Brown Instrument Division of Minneapolis-Honeywell Regulator Company in the way of suggestions, advice, original terminology, and constructive criticism.

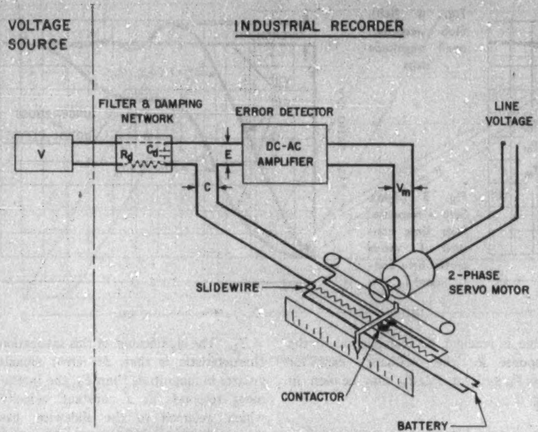


Fig. 1. Electronic self-balancing system

The system functions in such a manner that the command signal V, developed by a convenient voltage source, and the controlled signal C are equal under steady-state conditions, so that the error signal E is zero. Command signals are usually of the order of a few microvolts to 100 millivolts, depending on the electrical span of the voltage source. Electrical spans vary, but values of the order of 5 to 100 millivolts are typical. As a first step in the analysis, the system must be represented by a functional block diagram.

System Block Diagram

The presence of the R_dC_d network, Fig. 1, necessitates the preliminary evolution of a Laplace transform for the error signal, i.e., $E=F(s)$. By use of the principle of superposition and under the assumption that the amplifier impedance is resistive and much greater than slidewire resistance, it is demonstrated in the Appendix that

$$E(s)=\frac{A}{AT_1s+1}V(s)-\frac{A(T_1s+1)}{AT_1s+1}C(s) \qquad (1)$$

Thus, the equivalent block diagram of the voltage source and filter is that given in Fig. 2. Note particularly that a lead component is introduced in the feedback circuit.

Assume that the amplifier is a linear element having a frequency invariant transfer function

$$K_a=\frac{V_m}{E} \qquad (2)$$

and that the 2-phase induction motor[3] has a transfer function of the form

$$\frac{\theta_m(s)}{V_m(s)}=\frac{K_m}{S(T_ms+1)} \qquad (3)$$

By adding the equivalent blocks for the remaining control elements, i.e., amplifier, servomotor, and slidewire, to Fig. 2 the scheme of system operation is described by the block diagram shown in Fig. 3.

Transfer Function of Electronic Self-balancing System

It is demonstrated in the Appendix that the mathematical solution of the block diagram, Fig. 3, results in the system-and-error at amplifier transfer functions

$$\frac{\text{controlled voltage}}{\text{command voltage}}=G(s)=\frac{NA}{AT_1T_ms^3+(AT_1+T_m)s^2+(1+NAT_1)s+NA} \qquad (4)$$

$$\frac{\text{error voltage at amplifier}}{\text{command voltage}}=\frac{E(s)}{(Vs)}=\frac{As(T_ms+1)}{AT_1T_ms^3+(AT_1+T_m)s^2+(1+NAT_1)s+NA} \qquad (5)$$

Since any cubic expression is factorable into a first- and second-order term, the denominator is of the form $(T_zs+1)(s^2/\omega_1^2+2\zeta s/\omega_1+1)$ where T_z is an arbitrary designation for the first-order time contact, ω_1 is the conventional natural frequency, and ζ is the damping ratio. A particular form of the transfer function of interest occurs when $T_1=T_m$. Equations 4 and 5 then become equations 6 and 7

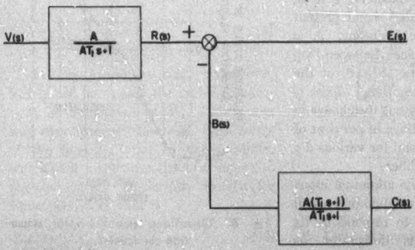
$$G(s)=\frac{1}{T_1s+1}\,\frac{1}{s^2/(N/T_1)+s/NA+1} \qquad (6)$$

$$\frac{E(s)}{V(s)}=\frac{s}{N}\,\frac{1}{s^2/(N/T_1)+s/NA+1} \qquad (7)$$

This is a condition which, if met by a system, leads to response without overshoot in a minimum of time.

Linear Step Response, Small Steps

As a simple illustration of the method of solution, consider the particular case



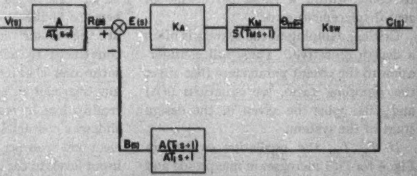


Fig. 2 (left). Voltage source and filter equivalent block diagram

Fig. 3 (above). Electronic self-balancing system equivalent block diagram

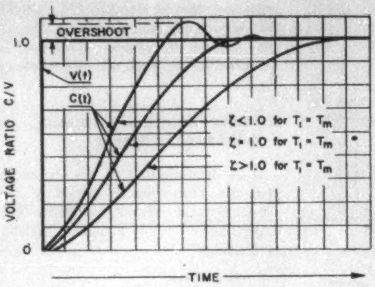


Fig. 4 (left). Step response, small magnitude steps

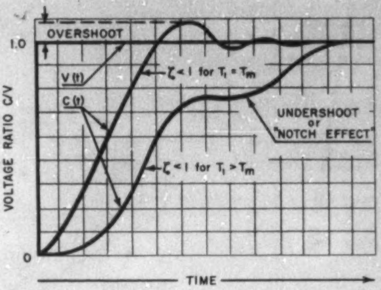


Fig. 5 (right). Step - response, filter time constant T versus motor time constant T_m

$T_1=T_m$ and the input signal is a step function, i.e.,

$$V(s)=\frac{V_e}{s} \qquad (8)$$

where V_e is the step magnitude in volts.

$$C(s)=\frac{V_e}{s}\,\frac{1}{T_1s+1}\,\frac{1}{s^2/\omega_1^2+2\zeta_1 s/\omega_1+1} \qquad (9A)$$

$$E(s)=\frac{V_e}{N}\,\frac{1}{s^2/\omega_1^2+2\zeta_1 s/\omega_1+1} \qquad (9B)$$

where

$$\zeta_1=\frac{1}{2A\sqrt{NT_1}} \text{ and } \omega_1=\sqrt{N/T_1}$$

The transient response of the system to small steps can be found by taking the inverse transform of equation 9. Three solutions to the equation are possible, since the damping ratio ζ may be less than, equal to, or greater than unity. To obtain the exact time solution, refer to any table of transforms.

Within the limits of linearity, typical system response is that shown in Fig. 4. The system can be underdamped for ζ_1 less than unity, in which case the response is characterized by overshoot; it can be critically damped in that the system reaches zero-displacement error within the smallest possible time without any overshoot; or it can be overdamped in that it takes a longer interval to come within a small percentage of zero-displacement error.

Critical damping of the system is often a design objective. Thus, full consideration to the circuit parameters that affect the damping ratio, see equations 9(A) and (B), must be given in the design stage of the system.

If $T_1<T_m$, the oscillation evident in Fig. 4 for $\zeta<1$ increases in magnitude and frequency, unless the system gain constant N is appreciably reduced. If $T_1>T_m$, the oscillations (for the oscillatory case) occur before the desired value is reached. A comparison of the response for the oscillatory case for $T_1=T_m$ and for $T_1>T_m$ may be seen in Fig. 5.

Component Saturation

In the system block diagram, Fig. 3, K_m and T_m are represented as constants and linear representation of the amplifier has been made. It is generally recognized that K_m and T_m are not constants but vary with the operating point on the torque-speed curve.[3] Actually, this departure from a constant value is less significant to system analysis than the fact that the servomotor saturates, i.e., it reaches a finite maximum or saturation velocity at some input to the motor control winding. Motor saturation before amplifier saturation is considered consistent with good design practice. This nonlinearity does result, however, in differences in system response to large changes in input level as compared with that to small changes in input level. Under these conditions, the conventional motor transfer function given previously would require modification by use of a describing function.[3]

In any event, whether a describing function or some other method of modification is used, the limiting input to the saturated element[4] must be known. In the case of the self-balancing system, because of the modulated signal present at the motor control winding, it is convenient to determine the open-loop saturation characteristic, Fig. 6, of the amplifier-motor combination. This is readily done by reading out the change in slidewire contactor voltage in per cent of electrical span per second for various d-c input levels to the amplifier.

The magnitude of the minimum input to the amplifier-motor combination which will cause the motor to respond at its maximum velocity is designated as $\pm E_L$. The significance of this saturation characteristic is that, for error signals greater in magnitude than E_L, the instrument responds at a constant velocity which, referred to the slidewire, has arbitrarily been called the saturation velocity and symbolized as $V_{v\text{-sat}}$. From this characteristic obtained under actual load, the system gain constant N can readily be determined by idealizing the curve, as shown in Fig. 6. Then N equals the slope of the curve multiplied by the slidewire voltage gradient. Thus N has the dimension of sec^{-1} as illustrated in the dimensional equation

$$\frac{1}{\text{second}}=\frac{\%\ \text{electrical span/second}}{\text{microvolts}}\times\frac{\text{microvolts}}{100\%\ \text{electrical span}}$$

Idealizing the curve presupposes that the dead band is of the order of a few microvolts at most. Dead band in itself is a nonlinearity which, because of its small relative magnitude, is considered of minor importance in the analysis of the dynamic response of the system to all but the very smallest signals.

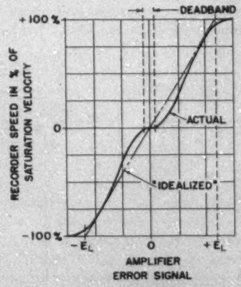


Fig. 6. Open-loop amplifier-motor saturation characteristic

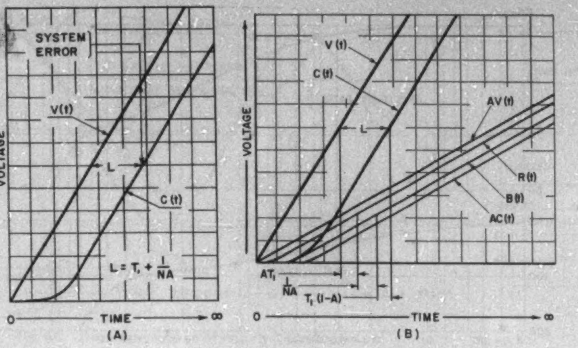


Fig. 7. Ramp response

A—Ramp response time L, lag time

B—Components of lag L

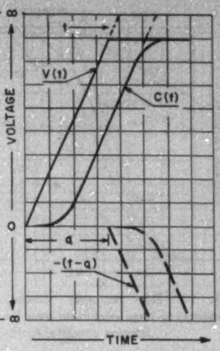


Fig. 8. Limited ramp response

Limit of Linear Step Response

In practice, the system saturates for small magnitude step changes in input level. As an illustration, consider again the particular case where $T_1 = T_m$ and where the input signal is a step function of magnitude V_c. Assume that the system is critically damped by substituting $\zeta_1 = 1$ in equation 9(B), and factor the resultant expression. The error signal thus becomes

$$E(s) = \frac{V_c}{N} \frac{1}{(s/\omega_1 + 1)^2} \tag{10}$$

where

$$\omega_1 = \sqrt{N/T_1}$$

The inverse transform, as obtained from a table[8] of transform pairs, is

$$E(t) = \frac{V_c}{N} \omega_1^2 t \, \epsilon^{-\omega_1 t} \tag{11}$$

The maximum value of the error voltage is then

$$E_{\max} = \frac{V_c}{N} \omega_1 \epsilon^{-1}$$

at time $t = 1/\omega_1$, or

$$E_{\max} = \epsilon^{-1} \frac{V_c}{\sqrt{NT_1}} \tag{12}$$

Let the minimum step magnitude, which will just cause the motor to saturate, be designated as V_{sat}. Then, the step magnitude E_L (see Fig. 6), in terms of the limiting input to the amplifier, which will just cause the instrument to respond at its maximum velocity for some portion of time is

$$V_{sat} = \epsilon \sqrt{NT_1} \, E_L \tag{13}$$

$N = 80$/second, $T_1 = 0.050$ second, and $E_L = 50$ microvolts are typical values for the system whose gain characteristic is given in Fig. 6. Substitution of these values in equation 12 reveals that the system would saturate for an input step of 0.272 millivolt. Compare this value with the electrical span of 5 millivolts for which the saturation characteristic of Fig. 6 was obtained:

$$\frac{V_{sat}}{V_{span}} = \frac{0.272 \text{ millivolt}}{5.000 \text{ millivolts}} = 5.4\%$$

It is evident that the system saturates or velocity limits for all but the smaller steps.

In a similar manner, it can be shown that damping ratios less than unity result in system saturation for input step signals of even smaller magnitude. For damping ratios in excess of unity, system saturation occurs at larger input signals.

Ramp Response

For the purposes of this paper, a ramp driving function is defined as a constant rate of change of the input level, i.e.,

$$V(t) = V_R t \tag{14}$$

where V_R is the ramp velocity in volts per second. From a table of transforms

$$V(s) = \frac{V_R}{s^2} \tag{15}$$

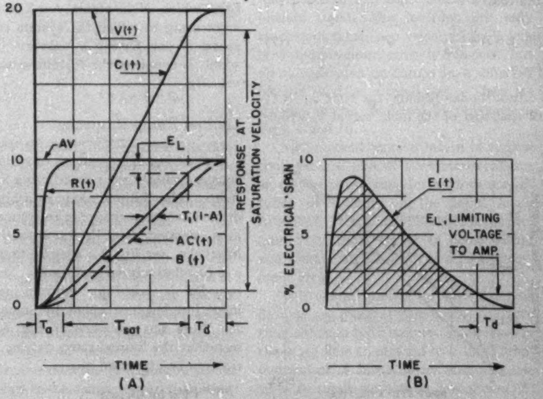


Fig. 9. Step response

A—Large magnitude steps

B—Amplifier error signal

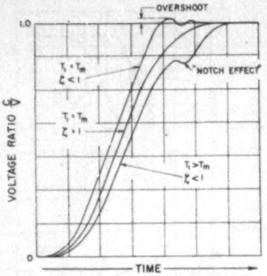


Fig. 10. Step response, comparison of responses to large magnitude steps for various damping conditions

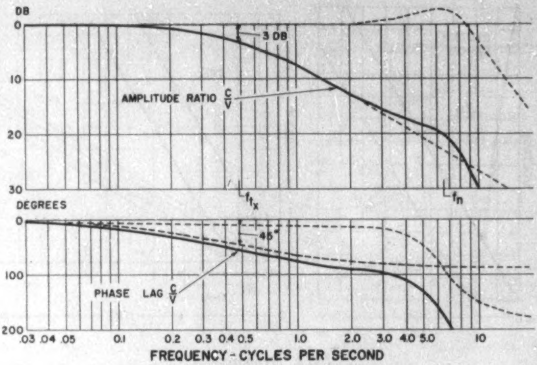


Fig. 12. Frequency response in linear region

The response to the ramp-driving function is obtained by taking the inverse transform of the response function

$$C(s) = \frac{V_R}{s^2} \frac{1}{T_x s + 1} \frac{1}{s^2/\omega_1^2 + 2\zeta s/\omega_1 + 1} \tag{16}$$

In the time domain, the response to a ramp has three possible solutions since the damping ratio may be less than, equal to, or greater than unity. One of the exact solutions is included in the Appendix.

Ramp Response Time (Lag Time)

In general, a linear instrument accelerates to the constant velocity of the input signal, but lags behind it by some time L, with a resultant system error as indicated in Fig. 7(A). By use of the final-value theorem, it is shown in the Appendix that

$$\lim_{t \to \infty} L = T_1 + \frac{1}{NA} \tag{17}$$

In a similar fashion, the time lag AT_1 of that part of the input signal V, which appears at the amplifier terminals, can be found. By extending the method, it can be shown that the feedback voltage leads the controlled voltage by the time $T_1(1-A)$. The separation of the individual lags is illustrated in Fig. 7(B). In addition, note that the amplifier actually sees only a part of the input signal and a part of the controlled voltage because of the voltage-divider effect of the amplifier and the damping impedance. This ratio is defined as A. Practically, the system lag L is a constant independent of the slope of the ramp only to the extent that N is constant. That L varies with the slope of the ramp is evident from a consideration of Fig. 6, saturation characteristic of the amplifier-motor combination, from which N is obtained. The maximum ramp to which the system can respond is one having a velocity (slope) which is equal to the system saturation velocity.

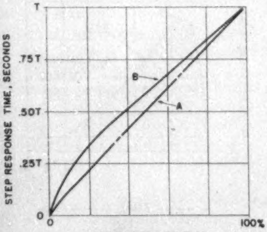


Fig. 11. Step-response characteristics

Limited Ramp Response

For use in a later analysis of system response, consider the response to a limited ramp. The limited ramp is arbitrarily defined here as a change in reference level at a constant rate V_R for a fixed time interval a; see Fig. 8. Assume that time a is of relatively large duration, i.e., the recorder is faithfully reproducing the input but lagging behind by some time L. From the geometry of Fig. 8 it is seen that the limited ramp driving function can be expressed as

$$V(t) + \begin{cases} V_R t \text{ for } 0 < t < a & \text{(18A)} \\ V_R\,[t-(t-a)] \text{ for } t > a & \text{(18B)} \end{cases}$$

The Laplace transform of equation 18(B) is given as equation 19

$$V(s) = \frac{V_R}{s^2} - \frac{V_R}{s^2} \epsilon^{-as} \tag{19}$$

where V_R is the slope of the ramp. For such a change in input level, the response of the system can be regarded as the response to the positive ramp, beginning at time $t=0$, less the response to the negative ramp beginning at time $t=a$, as illustrated in Fig. 8.

The steady-state voltage level is reached when the response to the negative ramp cancels that to the positive ramp. The system damping shapes the acceleration and deceleration regions of the response.

Nonlinear Step Response, Large Steps

The use of the graphical method of analysis is expedient in the examination of the large step-response characteristics

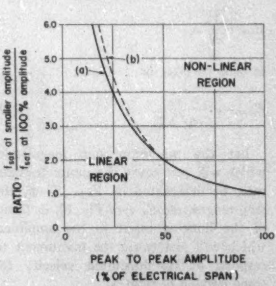


Fig. 13. Saturation barrier

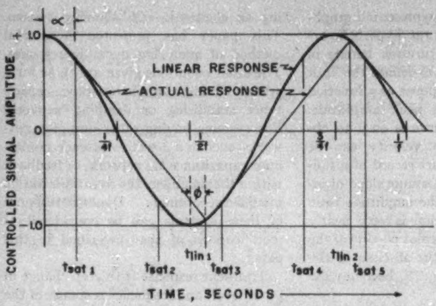


Fig. 14. Frequency response of velocity-limited system

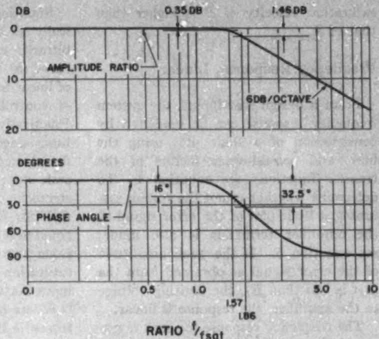


Fig. 15. Approximate frequency response for velocity-limited self-balancing systems

since system saturation (velocity limiting) occurs.

The transient error signal at the amplifier terminals (refer to Figs. 2 and 3, and equation 1) consists of the difference between the reference voltage $R(t)$ and the feedback voltage $B(t)$ at the summing point. These voltages are represented in their correct time relation in Figs. 9(A) and (B) for a system having a damping ratio ζ equal to unity.

For the step change in input level V, the forward loop component $R(t)$ of the error signal exhibits the wave form of the simple but relatively swift exponential growth of the filter condenser voltage to a new level AV. Initially, this change in R results in an error signal of sufficient magnitude to cause the servomotor to accelerate during time T_a to its saturation velocity.

During time T_{sat}, i.e., while the error signal is greater than the limiting voltage E_L to the amplifier, the servomotor continues to respond at its saturation velocity and the slidewire or controlled voltage varies accordingly. Compare Figs. 9(A) and (B). Note particularly that, during the recorder response at the saturation velocity, the feedback voltage $B(t)$ is a fractional part A of the controlled variable C (slidewire voltage) and leads it in time by an amount equal to $T_1(1-A)$.

Finally, the feedback signal reduces the error signal to the amplifier to some value less than the limiting voltage E_L and the motor decelerates. Its velocity decreases during time T_d until the amplifier error is zero and zero-displacement error is restored.

Examine Fig. 9(A) in the vicinity of time T_d. Compare this response with the response to the ramp illustrated in Figs. 7(A) and (B). Observe that the feedback voltage lags the controlled voltage in both instances by the same time. If the slope of the ramp in Fig. 7 were equal to the saturation velocity, the amplifier error signal for the ramp would become E_L.

Mathematical analysis of the response during the deceleration period of a large step can then be made by considering the solution for the equivalent limited ramp (applied at the summing point) which would produce the same response for the same amplifier error signal through the saturation velocity period T_{sat}. Then, the response function (see equation 59(A) in the Appendix) for time t greater than (T_a+T_{sat}) is

$$C(s)=\left[\frac{V_R A T_1}{s}+\frac{V_R}{s^2}\right]+$$
$$[l-e^{-as}]\frac{1}{(T_x s+1)(s^2/\omega_1{}^2+2\zeta s/\omega_1+1)}$$
$$\text{for } t>(t_a+t_{sat}) \qquad (20)$$

Compare this with the response function equation 9(A) for a small step. Since the response functions are not the same, the nature of the deceleration during the damping time must be different.

As a matter of fact, it is apparent from an examination of equation 20 that the system damps as if the driving function at the beginning of the damping period were the sum of a positive step, a positive ramp, a time-displaced negative step, and a time-displaced negative ramp. The time displacement a of equation 20 is considered to be much longer than that required by the system to accelerate to the saturation velocity during time T_a of Fig. 9(A). Consequently, response to the positive step and ramp is at a constant rate.

Evaluation of equation 20 reveals that the response during the damping time T_d is similar to that for small steps. It may appear as overshoot, undershoot (notch effect), or critical damping, depending on the system parameters. These are illustrated in Fig. 10.

Step-Response Characteristic, Practical Representation

Previous considerations pointed out the fact that the response of the electronic self-balancing system to a step change in input level is a complex function of the magnitude of the step. For a large amplitude signal, the recorder not only reaches its saturation velocity long before zero-displacement error is reached, but also decelerates or damps in a different fashion than for a small step. These responses can be compared by reference to Figs. 4 and 10.

One practical representation of system response is a plot of the time-to-balance or step-response time for various magnitudes of changes in the input level. Arbitrarily, the time to balance is that required for the recorder to come within and remain within 0.1% of zero-displacement error.

Such characteristics for two different designs are illustrated in Fig. 11. Note that although both require approximately the same time to respond to a 100% step, instrument A has a faster step-response time to small steps than instrument B. The obvious interpretation of these curves is that instrument A can accelerate and decelerate faster than B, but the saturation velocity of B is higher than that of A.

Frequency Response, Linear

From a design standpoint, the system frequency response can be predicted by construction of a Bode plot using the first- and second-order factors of the transfer function; see equation 4. To determine whether or not the system saturates, a Bode plot of the error signal at the amplifier terminals is also made; see equation 5. If the peak amplitude of the error signal as obtained from the plot is less than E_L, the limiting voltage to the amplifier, the response is linear.

The frequency response for such a system is illustrated in Fig. 12. Note that the Bode diagram is the sum of a first-order system, having a corner frequency $= 1/Tx$, and a second-order system having a natural frequency $= \omega_1/2\pi$ and a damping ratio ζ.

From a performance standpoint, obtaining and plotting the Bode diagram from test data is more valid. If the system were to remain linear for all magnitudes and frequencies of the driving function, one Bode plot would characterize the system. System saturation affects the frequency response of the system in the dominant way described in the following.

Saturation Frequency, Limit of Linear Frequency Response

Assume that the recorder is faithfully reproducing a sine wave. The maximum instantaneous velocity of any sine wave is the peak amplitude times the angular velocity as derived as follows.

$$C(t) = C \sin \omega t \tag{21}$$

where C is the peak amplitude of the controlled or recorded signal. Then

$$\frac{dC}{dt} = C\omega \cos \omega t \tag{22}$$

and

$$\left(\frac{dC}{dt}\right)_{\max} = C\omega \tag{23}$$

Equate the maximum velocity of the sinusoidal response to the maximum or saturation velocity V_{v-sat} of the recorder and an expression for the frequency beyond which saturation will just begin to distort the wave shape results.

$$C\omega_{sat} = V_{v-sat} \tag{24}$$

$$2\pi f C_{sat} = V_{v-sat} \tag{25}$$

$$f_{sat} = \frac{V_{v-sat}}{2\pi C} \tag{26}$$

Equation 26 can be represented graphically as a first-quadrant hyperbola arbitrarily called the saturation barrier in curve (*a*) Fig. 13. This defines the limit of linear frequency response as a function of controlled variable peak amplitude. Practically, such a curve is easy to obtain, since saturation velocity can be determined from a chart record of a full-scale step; V_{v-sat} = maximum slope of instrument record. If the amplitude ratio C/V for linear operation is very nearly equal to unity, as in the flat portion of the Bode plot in Fig. 12, the abscissa of the saturation barrier, Fig. 13, becomes the input amplitude V.

Where the amplitude ratio as obtained from the Bode plot, Fig. 12, differs from unity, the peak-to-peak value of the sinusoidal input signal which causes velocity saturation will change accordingly. As linear attenuation of the input signal by the filter generally occurs with higher frequency input signals, a greater amplitude can be applied without saturation, and the saturation barrier becomes that shown in curve (*b*) of Fig. 13.

Approximate Frequency Response Beyond Saturation Frequency

If the input signal for a given amplitude is varied at a frequency greater than the saturation frequency, distortion, attenuation, and phase shift result. The controlled or recorded wave form becomes progressively triangular in nature. An illustration of this nonlinearity is given in Fig. 14. Details are included in the Appendix.

It is shown in the Appendix that for frequencies lower than 1.86 f_{sat}, attenuation is less than 1.46 decibels. Beyond this frequency, the system attenuates because of velocity limiting at a rate of approximately 6 decibels per octave. This information is plotted in Fig. 15, which gives the approximate frequency response for velocity-limited self-balancing systems. Since the abscissa of the curve is the ratio of frequency of the input to the saturation frequency, the plot is valid for any signal amplitude if the saturation velocity is known. Frequency response for large amplitude signals thus can be represented on a Bode plot as a family of curves.

Summary

An error detector, a slidewire and its associated bridge voltage, a positioning servomotor, and means for damping the motor speed are the fundamental components of most industrial recorders utilizing an electronic self-balancing system. This paper has presented a general method of analyzing dynamic response. The exact equations given are those for a system having a simple resistance-capacitance stabilizing or damping network. For systems having other damping networks, such as a 3-section cascaded resistance-capacitance lag network, or feedback within the amplifier, the equations can be modified accordingly. Dynamic response of these systems can be considered as modifications of that described in this paper.

Transient response (to a step change in input level) and frequency response of the system were mathematically analyzed. Departure of the system from the linear condition was evaluated in terms of the dominant nonlinearity, the saturation velocity. It was demonstrated that system response to a large step differs considerably from that to a small step because of velocity saturation. The effect of damping on the response time for a step change in input level was also shown. The influence of this dominant nonlinearity on frequency response was analyzed and practically illustrated in a characteristic called the saturation barrier. It is shown that this barrier can readily be approximated for the industrial recorder from a knowledge of the saturation velocity.

Appendix

In the following derivations, use is made of existing charts of stabilizing-network transfer functions and tables of transform pairs, where possible. The source of each chart or transform pair is identified in the references.

Derivation of Block-Diagram Equivalent Circuit of Voltage Source and Filter Elements

Assume that the input impedance of the amplifier is resistive and that the slidewire resistance is small enough to be considered negligible. The simplified block diagram of the system is given in Fig. 16(A). Applying the principle of superposition, the forward-loop and the feedback components E_1 and E_2, respectively, of the amplifier error signal are derived from Figs. 16(B) and (C), where the forward-loop component transfer function is obtained from a chart[3] of stabilizing-network transfer functions. It is

$$\frac{E_1(s)}{V(s)} = \frac{A}{AT_1s+1} \tag{27}$$

The feedback-component transfer function as obtained from the same source is

$$\frac{E_2(s)}{C(s)} = \frac{A(T_1s+1)}{AT_1s+1} \tag{28}$$

By the principle of superposition, for negative feedback

$$E(s)=E_1(s)-E_2(s) \tag{29}$$

$$E(s)=\frac{A}{AT_1s+1}\,V(s)-\frac{A(T_1s+1)}{AT_1s+1}C(s) \tag{30}$$

Equation 30 appears in diagram form in Fig. 2.

Derivation of System Transfer Function

From the system block diagram, shown in Fig. 3, the transfer function is

$$\frac{C(s)}{R(s)}=\frac{\dfrac{K_aK_mK_{sw}}{s(T_ms+1)}}{1+\dfrac{A(T_1s+1)}{AT_1s+1}\,\dfrac{K_aK_mK_{sw}}{s(T_ms+1)}} \tag{31}$$

Equation 31 is reduced to equation 32 where $N=K_aK_mK_{sw}$, as

$$\frac{C(s)}{R(s)}=\frac{N(AT_1s+1)}{AT_1T_ms^3+(AT_1+T_m)s^2+(1+NAT_1)s+NA} \tag{32}$$

But, as determined from the system block diagram, Fig. 3, the reference voltage R is related to the command voltage V

$$R(s)=\frac{A}{AT_1s+1}V(s) \tag{33}$$

Thus, from equations 32 and 33 the closed-loop system transfer function becomes the following

$$\frac{C(s)}{V(s)}=\frac{NA}{AT_1T_ms^3+(AT_1+T_m)s^2+(1+NAT_1)s+NA} \tag{34}$$

Derivation of Error-Voltage Transfer Function

From the block diagram shown in Fig. 3

$$\frac{C(s)}{E(s)}=\frac{N}{s(T_ms+1)} \tag{35}$$

$$E(s)=\frac{s(T_ms+1)}{N}\,C(s) \tag{36}$$

$$\frac{E(s)}{V(s)}=\frac{s(T_ms+1)}{N}\,\frac{C(s)}{V(s)} \tag{37}$$

From equations 34 and 37, then

$$\frac{E(s)}{V(s)}=\frac{s(T_ms+1)}{N}\times\frac{NA}{AT_1T_ms^3+(AT_1+T_m)s^2+(1+NAT_1)s+NA} \tag{38}$$

Derivation of Ramp Response Time L and Components of Lag

Refer to Fig. 7(A) and let

$$V(t)=V_Rt \tag{39}$$

where V_R is the constant velocity of the ramp in volts/second

$$V(s)=\frac{V_R}{s^2} \tag{40}$$

$$C(s)=\frac{NA}{AT_1T_ms^3+\ldots+NA}\,\frac{V_R}{s^2} \tag{41}$$

The system error is

$$V(s)-C(s)=\frac{V_R}{s^2}-\frac{NA}{AT_1T_ms^3+\ldots+NA}\,\frac{V_R}{s^2} \tag{42}$$

$$V(s)-C(s)=\frac{V_R}{s^2}\times\frac{AT_1T_ms^3+(AT_1+T_m)s^2+(1+NAT_1)s}{AT_1T_ms^3+\ldots+NA} \tag{43}$$

From the final-value theorem

$$\lim_{t\to\infty} f(t)=\lim_{s\to 0} s\mathrm{F}(s) \tag{44}$$

$$\lim_{t\to\infty}[V(t)-C(t)]=\lim_{s\to 0} V_R\times\frac{AT_1T_ms^2+(AT_1+T_m)s+(1+NAT_1)}{AT_1T_ms^3+\ldots+(1+NAT_1)s+NA} \tag{45}$$

$$\lim_{t\to\infty}[V(t)-C(t)]=V_R\,(T_1+1/NA) \tag{46}$$

But

$$\frac{\text{displacement (volts)}}{\text{velocity (volts/sec)}}=\text{time in seconds} \tag{47}$$

Then

$$\frac{\lim[V(t)-C(t)]}{V_R}=L \tag{48}$$

$$\text{lag time } L=T_1+1/NA \tag{49}$$

from equations 46 and 48.

In a similar manner, it can be demonstrated that the feedback voltage $B(t)$ lags the input ramp $V(t)$ by an amount less than the system lag. Subtracting the system lag from this value reveals that the feedback voltage $B(t)$ leads the controlled voltage $C(t)$ by $T_1(1-A)$ seconds. This is illustrated in Fig. 7(B).

Derivation of System Response to Ramp

Using the factored form of the transfer function, the response to a ramp of unity velocity is given by

$$C(s)=\frac{1}{s^2}\,\frac{1}{T_xs+1}\,\frac{1}{s^2/\omega_1^2+2\zeta s/\omega_1+1} \tag{50}$$

Although three solutions to equation 50 are possible, consider, for illustration purposes, the case where $\zeta>1$. The second-order term in the denominator of equation 50 is factorable and the response function becomes

$$C(s)=\frac{1}{s^2}\,\frac{1}{T_xs+1}\,\frac{1}{T_ys+1}\,\frac{1}{T_zs+1} \tag{51}$$

where T_y and T_z are arbitrary designations for roots of the second-order equation. Then

$$C(t)=t-K_1-K_2e^{-t/T_x}-K_3e^{-t/T_y}-K_4e^{-t/T_z} \tag{52}$$

where K_1, K_2, K_3, and K_4 are constants, dependent upon T_x, T_y, and T_z. For the exact relation see reference 5, equation 00.230. For these purposes, observe that the response is a ramp function less that of an exponential growth function. The general nature of the response is illustrated in Fig. 7(A).

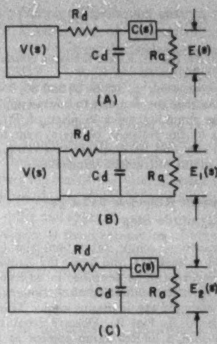


Fig. 16. Simplified block diagram of amplifier error system

A—Basic diagram
B—Forward-loop component
C—Feedback component

Derivation of Response to Ramp Applied at Summing Point

For use in analyzing the response to step input signals of large magnitude, given later, assume that an equivalent ramp driving function could be applied in place of the voltage $R(t)$ in Fig. 3. Factoring the denominator of equation 32 obtains

$$\frac{C(s)}{R(s)}=\frac{AT_1s+1}{A}\,\frac{1}{T_xs+1}\,\frac{1}{s^2/\omega_1^2+2\zeta s/\omega_1+1} \tag{53}$$

Assume that the damping ratio is greater than unity and $R(s)=AV_R$, a ramp having a slope equal to the instrument saturation velocity. Then

$$C(s)=\frac{AV_R}{s^2}\,\frac{AT_1s+1}{A}\,\frac{1}{T_xs+1}\times\frac{1}{(T_ys+1)(T_zs+1)} \tag{54}$$

where T_y and T_z are again arbitrary designations for roots of the second-order term in the transfer function. Solve equation 54 as a sum of partial fractions. Finding the inverse transforms from a convenient table

$$C(s)=V_R\frac{AT_1}{s}\,\frac{1}{T_xs+1}\,\frac{1}{T_ys+1}\,\frac{1}{T_zs+1}+\frac{V_R}{s^2}\,\frac{1}{T_xs+1}\,\frac{1}{T_ys+1}\,\frac{1}{T_zs+1} \tag{55}$$

$$C(t)=AV_RT_1\left(1-\frac{K_2}{T_x}e^{-t/T_x}-\frac{K_3}{T_y}e^{-t/T_y}-\frac{K_4}{T_z}e^{-t/T_z}\right)+V_R\Big(t-K_1+$$

$$K_2\epsilon^{-t/T_2}+K_3\epsilon^{-t/T_3}+K_4\epsilon^{-t/T_4}\Big) \qquad (56)$$

where K_1, K_2, K_3, K_4 are the same constants as in equation 52.

In comparing equations 52 and 56, it can be seen that the response to a ramp applied at the summing point in place of $R(t)$ is equal to the response of the system to a small step and the response to a ramp applied at the instrument terminals.

Transient Response to Large Magnitude Step

Because of motor saturation, the controlled voltage $C(t)$ and hence the feedback voltage $B(t)$ can be described in terms of a ramp soon after initiation of a large magnitude step change in input level. This is illustrated in Fig. 9. Observe from this figure that a limited ramp applied at the summing point in place of the voltage $R(t)$ would describe the minimum equivalent reference signal which would result in response at a constant velocity and, consequently, a constant error signal at the amplifier terminals, prior to the initiation of damping. (The limited ramp could be shown in Fig. 9 as a voltage just greater than $B(t)$ by a constant voltage E_L during time t_{sat}.) To describe the nature of damping, let the limited ramp become

$$R(s)=\frac{AV_R}{s^2}-\frac{AV_R}{s^2}\epsilon^{-as}=\frac{AV_R}{s^2}(1-\epsilon^{-as}) \qquad (57)$$

In equation 57, V_R is the slope of a ramp equal to the system saturation velocity $V_{s\text{-}sat}$ in volts/sec. Since response equals the driving function times the transfer function, the following equation may be obtained from equations 32 and 57

$$C(s)=\frac{AV_R}{s^2}(1-\epsilon^{-as})\times \frac{(AT_1s+1)}{A}\,\frac{NA}{AT_1T_ms^3+\ldots+NA} \qquad (58A)$$

From which

$$C(s)=\left[\frac{V_RAT_1}{s}+\frac{V_R}{s^2}\right](1-\epsilon^{-as})\times \frac{NA}{AT_1T_ms^3+\ldots+NA} \qquad (58B)$$

Since the last factor in equation 58(B) is the system transfer function, the response to a limited ramp applied at the summing point is equivalent to that for a ramp and a step applied at the instrument terminals less the time-displaced negative equivalents of the ramp and the step.

Let the system transfer function, as it appears in equation 58(B) take on its factored form; see equations 5 through 7. Then

$$C(s)\times\left[\frac{V_RAT_1}{s}+\frac{V_R}{s^2}\right](1-\epsilon^{-as})\times \frac{1}{(T_xs+1)}\,\frac{1}{(s^2/\omega_1^2+2\zeta s/\omega_1+1)} \quad \text{for } t>t_{sat} \qquad (59A)$$

For the exact-time solution, equation 59(A) can be evaluated for time $t>t_{sat}$ by partial fractions

$$C(t)=\mathfrak{L}^{-1}\left[\frac{AV_R}{s^2}\,\frac{AT_1s+1}{A}\,G(s)\right]-\mathfrak{L}^{-1}\left[\frac{AV_R}{s^2}\,\frac{AT_1s+1}{A}\,\epsilon^{-as}\,G(s)\right] \qquad (59B)$$

where $G(s)$ is defined in equation 4.

Previous to time t_d, the system was assumed to have been responding at the saturation velocity. Graphical solution of the positive inverse transform of equation 59(B) for $t>t_{sat}$ can expediently be made by simply extending the $C(t)$ line of Fig. 9(A) through time t_d. From this plot, the time equivalents of the negative inverse transform indicated in equation 59(B) can be subtracted to obtain the exact response during the damping periods. Closer examination of equations 59(A) and (B) indicates that response to large magnitude steps during damping time t_d is similar to that for small steps. Overshoot, undershoot or even critical damping may appear.

Approximate Frequency Response Beyond Saturation Frequency

Analyze Fig. 13 for the range of frequencies $f_{sat}<f<1.86f_{sat}$, assuming that initially there is no error in the output because of motor saturation, as from time $t=0$ to time t_{sat1}. At time t_{sat1} the saturation velocity is just reached. The value of t_{sat1} is obtained by equating the saturation velocity, equation 24, to the instantaneous velocity of the output wave.

$$C\omega_{sat}=C\omega\sin\omega t_{sat} \qquad (60)$$

$$\sin\omega t_{sat}=\frac{\omega_{sat}}{\omega}=\frac{f_{sat}}{f} \qquad (61)$$

$$t_{sat}=\frac{1}{2\pi f}\sin^{-1}f_{sat}/f \qquad (62)$$

From time t_{sat1} until the error between linear response and actual response decreases to zero, at t_{lin1}, the recorder can only respond at its maximum velocity. From time t_{lin1} the recorder responds linearly until its velocity limits the response again, at t_{sat2}. Response from t_{sat2} to t_{lin2}, can be analyzed in similar fashion. After this, the cycle is repeated. Due to acceleration limitations, the actual output may overshoot or undershoot.

It can be shown from the geometry of Fig. 14 that, at the frequency of 1.86 times the saturation frequency, t_{lin1} and t_{sat2} are equal, so that the wave form (neglecting acceleration) is theoretically completely triangular in shape. For frequencies greater than $1.86f_{sat1}$, the assumption is made that the output is periodic but that it has a sawtoothed wave form. It can be shown geometrically that such a wave form for a given sinusoidal input has a peak amplitude that is inversely proportional to the frequency. Fig. 15, which gives the approximate frequency response beyond saturation frequency, was derived from the previous analysis.

Sample Calculations for Fig. 15

The angle at which the system returns to the linear condition from the saturated velocity is obtained by equating the expression for the amplitude of the linear curve of Fig. 14 to that for the actual response at time t_{lin1}. From the geometry of Fig. 14, then

$$\cos(\pi+\phi)=\cos\alpha-\omega_{sat}\frac{\pi+\phi-\alpha}{\omega} \qquad (63)$$

Example 1, $f=\pi/2f_{sat}$

From equation 62 saturation occurs at time

$$t_{sat}=\frac{1}{2\pi f}\sin^{-1}\frac{2}{\pi}$$

or in angular degrees as

$\alpha_{sat}=39.50$ degrees $=0.69$ radian

Substitution in equation 63 gives the following equation

$$\cos(\pi+\phi)-\cos 39.5-\frac{2}{\pi}(\pi+\phi-.69)$$

$$\cos\phi=0.787+0.636\phi$$

From this equation, by trial and error, $\phi=16$ degrees. The amplitude ratio in decibels is

$$20\log\cos\phi=0.35$$

Example 2, $f=1.86f_{sat}$

From equation 62, saturation occurs at time

$$t_{sat}=\frac{1}{2\pi f}\sin^{-1}\frac{1}{1.86}$$

or in angular degrees as

$\alpha_{sat}=32.50$ degrees $=0.57$ radians

Substitution in equation 63 gives the following equation

$$\cos(\pi+\phi)=\cos 32.5-\frac{1}{1.86}(\pi+\phi-0.57)$$

$$\cos\phi=.539+.538\phi$$

From this equation, by trial and error, $\phi=32.5$ degrees. The amplitude ratio in decibels is

$$20\log\cos\phi=1.46$$

References

1. SERVOMECHANISMS AND REGULATING SYSTEM DESIGN, Vol. I (book), H. Chestnut. R. W. Mayer. John Wiley & Sons, Inc., New York, N.Y., 1951, pp. 194, 208–12.

2. PRINCIPLES OF SERVOMECHANISMS (book) C. S. Brown, D. P. Campbell. John Wiley & Sons, Inc., 1948, pp. 131–33.

3. SERVOMECHANISMS AND REGULATING SYSTEM DESIGN, Vol. II (book), H. Chestnut, R. W. Mayer. John Wiley & Sons, Inc., 1955, pp. 224, 225, 241–50; p. 126, Network no. 4. 2–3(a); p. 122, Network no. 4. 2-1(c).

4. APPROXIMATE FREQUENCY RESPONSE METHODS FOR REPRESENTING SATURATION AND DEAD BAND, H. Chestnut. *Transactions*, American Society of Mechanical Engineers, New York, N.Y., vol. 76, Nov. 1954.

5. PRINCIPLES OF AUTOMATIC CONTROLS (book), F. E. Nixon. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1953, pp. 373, 377, 380 and 397—equation nos. 00.020, 00.130, 00.230, 11.020.

Multiplex Telegraph Equipment for Radio and Submarine Cable Circuits

B. S. DIAMOND
MEMBER AIEE

TIME-DIVISION multiplex has provided the most efficient utilization of trunk telegraph circuits and, over the decades that have passed since Baudot developed the first really practical method, inventors and engineers have devoted a great deal of thought and research to the problem of increasing, the efficiency, effectiveness, and scope of multiplex operation.

In recent years, two basic improvements in long-distance, overseas telegraph circuits have stimulated further developments in time-division multiplex and some noteworthy improvements in techniques have been developed in both Europe and America. The first is that there has been a great improvement in reliability and continuity of radio circuits that has resulted from the general adoption of frequency-shift keying methods.[2–5] The second is that the rate of transmission on long submarine cables has been quadrupled in many cases by the insertion of submersible amplifiers offshore and that there is considerable promise of far greater increases in transmission rates in the not-too-distant future.[6–9]

The purpose of this paper is to review the transmission problems of both radio and cables from the point of view of a multiplex engineer and to describe methods developed by engineers of American Cable and Radio Corporation to enable advantage to be taken of improved transmission facilities.

Special Terms

Transmission, as used in this paper, covers the entire process of conveying signals from a sending terminal to a receiving terminal. It includes passage through one or more media, translation into the forms required in various media, amplification, regeneration, etc.

Bit is the unit signalling element.

Bauds refer to the rate of transmission in bits per second.

Two-condition operation is the common binary method of signaling in which line pulses are either marking or spacing.

Three-condition operation is the ternary method of operation, normally used when receiving signals on long submarine telegraph cables, in which line pulses may be one of three conditions.

Basic frequency, in normal operation, is the frequency at which alternating current is transmitted when alternate bits are of opposite polarity. On long submarine cables, where singles are suppressed and reception is effectively 3-condition, the basic frequency is one half of the maximum frequency at which bits may be transmitted.[8–10]

Distortion. If signals were all ideal, all bits, or groups of bits of the one polarity, would conform to their theoretical durations with transitions between polarities occurring instantaneously at intervals equal to one bit, or to the sum of the durations of an integral number of bits. Telegraph distortion may be defined as deviation from the ideal signal and is usually defined as percentage displacement. Distortion on telegraph signals may be divided into three types:

1. Characteristic distortion, which is a function of the transmission line with its sending and receiving equipment.
2. Fortuitous distortion, which results from random disturbances.
3. Bias distortion, in which the significant intervals are generally greater than the ideal on one side (space or mark as the case may be) and less than the ideal on the other.

Sway. This term is applied, particularly in the operation of submarine cables, to a transient bias.

Time-division multiplex is the process of dividing a given circuit into channels by the use of time intervals for channel separation. The simplest forms are

1. Where the circuit is switched between two channels on alternate bits. This is sometimes called interleaving.
2. Where the circuit is switched between two channels on alternate groups of bits. This is sometimes referred to as sequential distribution.

Bichannel or double refers to a multiplex system of two channels.

Triple refers to a multiplex system of three channels.

Quad or quadruple is a multiplex system of four channels.

Supermultiplex is a multiplex system with a complex distribution pattern and which is composed of "submultiplexes" that may be bichannels, triples, or quads.

Twinplex is a method of signaling whereby a carrier is shifted between four frequencies, only one of which is present at a time, in order to obtain two circuits over a long-distance radio beam. The four frequencies represent the four possible keying conditions, that is, mark and mark, mark and space, space and mark, and space and space.[10]

Polyplex is the combination of two, or more, twinplex transmissions, closely spaced as to frequency values. Thus, a 4-circuit polyplex system consists of two twinplex carriers, each shifted among its own four frequencies. The two carriers may be passed through a single linear amplifier or through two, well-decoupled nonlinear amplifiers. The specific values of frequency shift as well as frequency separation between the "center-frequency values" of the two carriers is selected so that the frequency values of any cross-modulation products between the two carriers will fall between, or outside of, selection filter passbands.[11]

Transmission Features of High-Frequency Radio Circuits

When a high-frequency radio signal is picked up at a remote receiving station, hundreds or thousands of miles from the sending station, reception is dependent on a complex process of refractions and reflections.

The transmitted beam is first refracted in the ionized layers of the upper atmosphere and returned to the earth's surface where it may be reflected back to the upper atmosphere for a second "hop." This may be repeated for an indefinite number of hops.

As the ionosphere is seldom stable and as the earth's surface is irregular, the transmitted beam is subject to diffusion and it is divided into an indefinitely large number of rays, each of which travels along a path of its own. The paths may

---

Paper 57-64, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 19, 1956; made available for printing January 8, 1957.

B. S. Diamond is with the American Cable & Radio Corporation, New York, N. Y.

The development of the equipment described in this paper was the result of team work involving many persons; however, most of the senior engineers are indebted to W. C. Peterman, now retired, for a large part of their knowledge of multiplex techniques.[1]

The assistance of George Royden in preparing this paper is gratefully acknowledged.

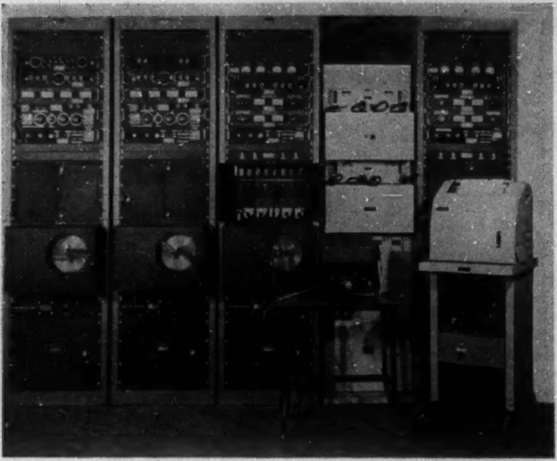


Fig. 1. Multiplex terminal telegraph equipment with monitoring teleprinters

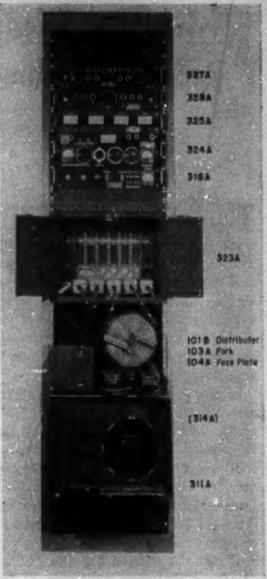

327A
326A
325A
324A
316A
323A
101B Distributor
103A Fork
104A Face Plate
(314A)
311A

Fig. 4 (right). Multiplex receiving rack type 320A

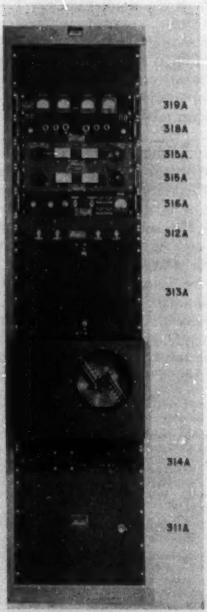

319A
318A
315A
315A
316A
312A
313A
314A
311A

Fig. 2. Multiplex transmitting rack type 310A

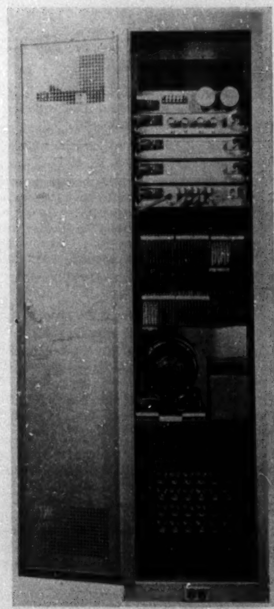


Fig. 3. Multiplex transmitting rack type 310A, rear view

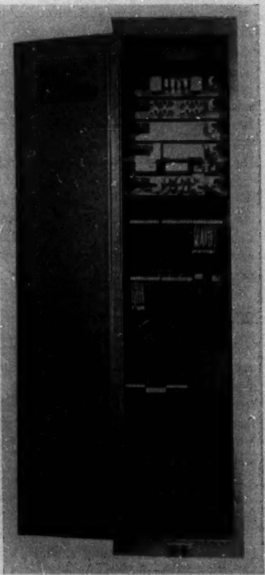


Fig. 5. Multiplex receiving rack type 320A, rear view

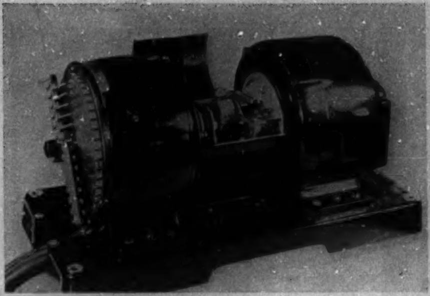


Fig. 6 (above). Multiplex distributor type 101B

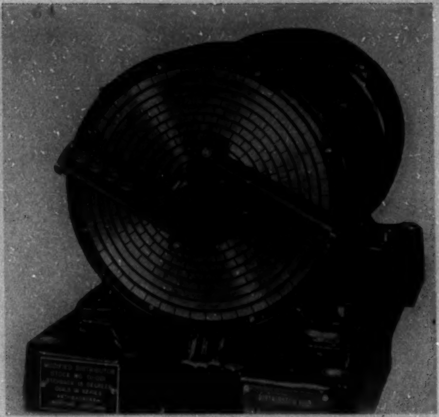


Fig. 7 (right). Face-plate type 104A and brush-holder type 102A

differ in length from one another and each is subject to continuous variation in path length and intensity.

A number of such rays, falling on a receiving antenna, will be integrated and may form a recognizable signal. Due to the continuous variation in phase and amplitude of the component rays, the strength of the received signal will wax and wane according to the vectorial summation of its components.

A second antenna, in a different location than the first, will pick up a signal similar to that of the first, however, the variation in amplitude due to the effects of multipaths will have a different pattern with the respective minima and maxima seldom coinciding. It is probable, therefore, that, at any one instant, at least one antenna will produce a usable signal. Hence continuity may be improved by circuitry which selects the stronger of the two signals or which produces a combined output toward which the contribution of each receiver is proportional to the square of the signal amplitude.

Since the height of the refracting layers and the number of hops taken by the strongest signal are not constant, the effective path of the received signal is subject to variation. As transit time is equal to distance divided by velocity, the transit time of the demodulated telegraph signals will vary with changes in effective length.

On long-distance high-frequency circuits, multipath effects place an upper limit on the rate multiplex signals can be received. There are also practical difficulties in the application of normal frequency-division methods. However, time-division multiplex can be applied to twinplex and polyplex radio transmissions and, if the derived circuits of twinplex or polyplex are operated at identical rates of speed, with synchronized transitions, the rate at which telegraph signals can be

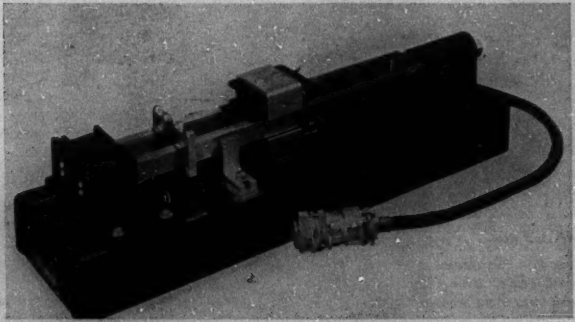


Fig. 8. Electronic fork type 103A

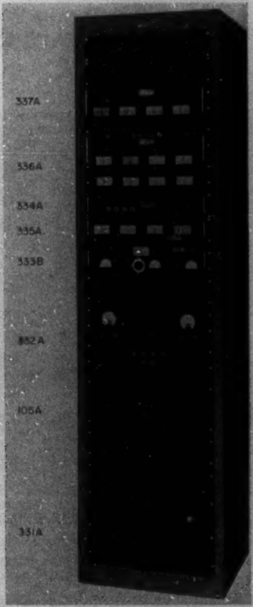


Fig. 9. Multiplex repeater rack type 330A

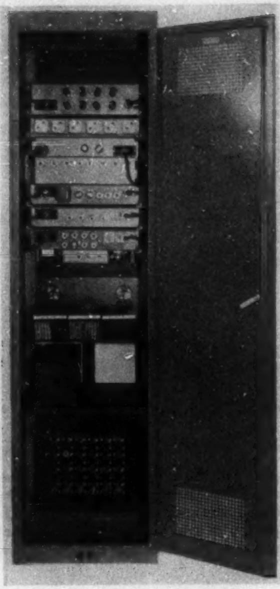


Fig. 10. Multiplex repeater rack, rear view

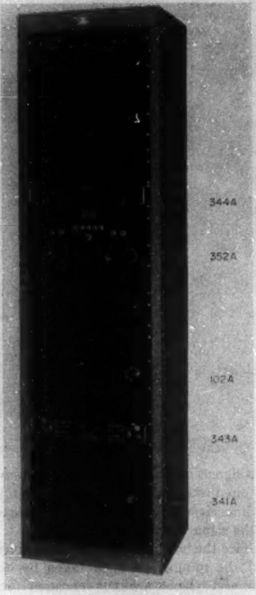


Fig. 11. Monitor satellite rack type 350A

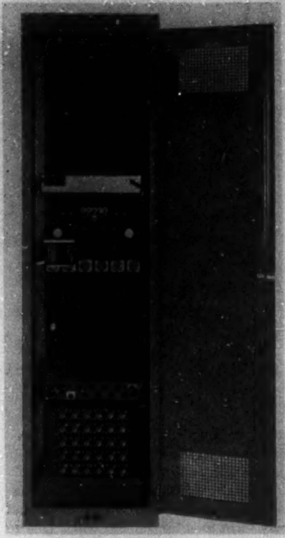


Fig. 12. Monitor satellite rack type 350A, rear view

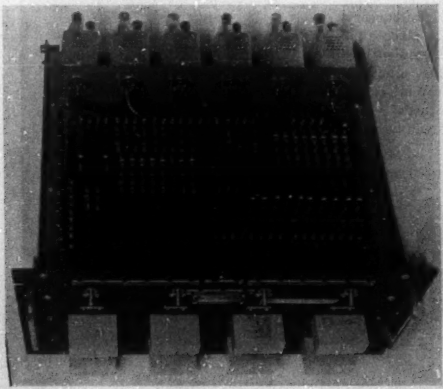


Fig. 13 (above). Distributing matrix type 336A, top view

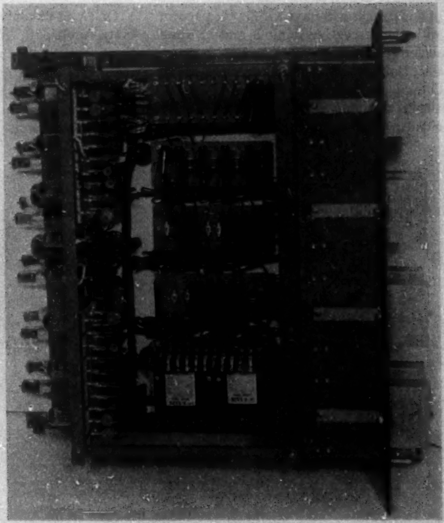


Fig. 14 (right). Distributing matrix type 336A, bottom view

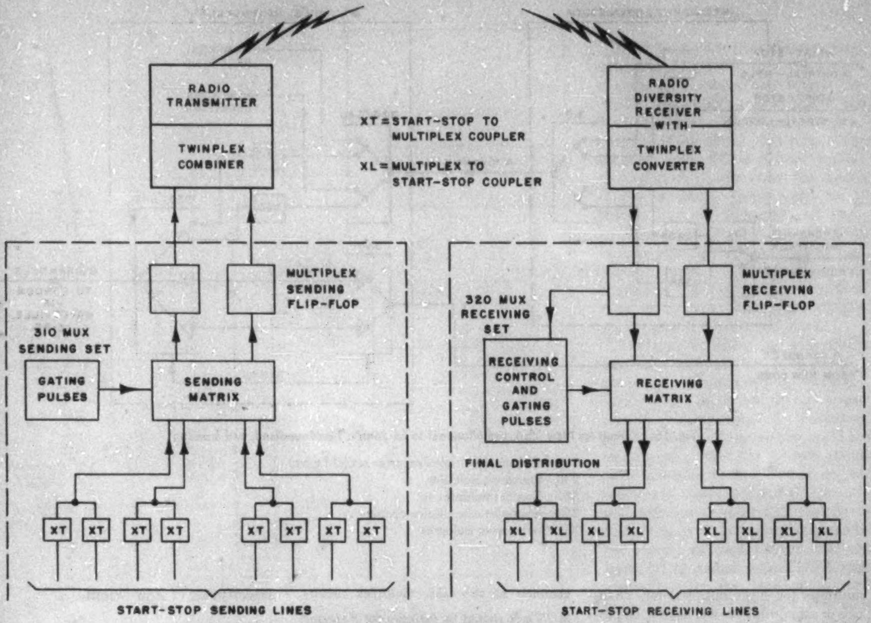

RADIO TRANSMITTER
TWINPLEX COMBINER
XT = START-STOP TO MULTIPLEX COUPLER
XL = MULTIPLEX TO START-STOP COUPLER
RADIO DIVERSITY RECEIVER WITH
TWINPLEX CONVERTER
MULTIPLEX SENDING FLIP-FLOP
310 MUX SENDING SET
GATING PULSES
SENDING MATRIX
320 MUX RECEIVING SET
MULTIPLEX RECEIVING FLIP-FLOP
RECEIVING CONTROL AND GATING PULSES
RECEIVING MATRIX
FINAL DISTRIBUTION
XT
XT
XT
XT
XT
XT
XT
XT
XL
XL
XL
XL
XL
XL
XL
XL
START-STOP SENDING LINES
START-STOP RECEIVING LINES

Fig. 15. Eight-channel operation over a twinplex radio circuit

transmitted over a long-distance radio circuit can be greatly increased.

Hence for optimum operation over long-distance beams with nonstable paths, sending multiplex equipment should provide for synchronized transmission over twinplex or polyplex circuits. On the receiving side, an efficient telegraph signal regenerator, or the equivalent, will be necessary and there should be means of maintaining synchronism that will enable phase shifting to be followed at a reasonable rate so as to maintain best tolerance for signal distortion and transit-time variation.

Transmission Features of Submarine Cable Circuits

In contrast with radio circuits, long submarine cables provide a medium which is extremely stable but with definite speed limitations due to the rise of attenuation with increase of signaling frequency. Unless loaded or repeatered, the limiting attenuation of, say, 80 or 90 decibels is reached at a low subaudio frequency. Without repeaters, the maximum basic receiving frequency for main cable sections operated by the company varies between a low of 5 cycles and a high of 18 cycles per second.

The high capital cost and limited spectrum of submarine cables have resulted in the development of techniques for working through heavy distortion. Long submarine cable circuits are considered as being good if the distortion proportion does not exceed 50 per cent. Many cables are operated, quite satisfactorily, with less than 50 per cent margin.

During the past decade, vacuum-tube amplifiers or repeaters have been developed with suitable characteristics for installation some distance off shore where the signal is slightly stronger and the noise level much lower. The theory of these repeaters and their application has previously been described.[6–10]

During the past few years, repeaters have been inserted in all but one of the company's main transatlantic cable sections and maximum speeds (in one direction) have been increased more than four times. Further increases are anticipated with the insertion of improved repeaters and by the use of tandem repeaters. With extensions of techniques already pioneered, rates of speeds of several hundreds of bauds can be anticipated on existing transatlantic cables within the foreseeable future.

The multiplex sets in use prior to repeaterization were not suitable for the higher rates of speed and greater number of channels that repeaters make possible. Furthermore, the important traffic centers at New York, N. Y.; London, England; Paris, France; Montreal, Que., Canada; etc., are remote from the main cable sections which run between Newfoundland and Nova Scotia in Canada on the one side, and Ireland on the other. The available overland and cable connections between the traffic centers and the terminals of the main cable sections cannot, in all cases, now be made as fast as the main sections; this is particularly so when such connections have to be established over carrier channels with maximum working speeds of 50 or 60 bauds. Hence, in developing a multiplex system suitable for repeatered cables, it is necessary to assume that means of splitting a main

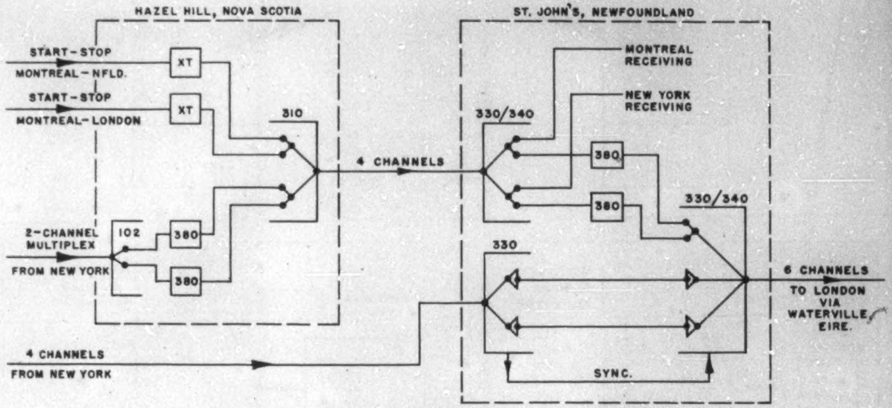


Fig. 16. Circuit for New York and Montreal to St. John's, Newfoundland, and London

102—2-channel multiplex repeater set (old type)
330—Electronic multiplex
340—Satellite multiplex set
360—Multiplex-to-multiplex coupler
XT—Start-stop to multiplex

multiplex (or supermultiplex) into channels or submultiplexes will have to be provided.

Design Criteria

With both cable and radio facilities expanding rapidly, the company required equipment suitable for the following applications:

1. Long cable chains of several links; particularly between New York and London with intermediate stations at Nova Scotia, Newfoundland, the Azores, and Ireland.
2. Point-to-point radio circuits which might be either straight frequency-shift, twinplex, or polyplex.
3. Chain radio circuits of two or more links.
4. Mixed cable and radio chains.

After considerable study, it was decided that the most practicable basis would be to retain the traditional face-plate methods on relatively low-speed sequential multiplexes of two channels which could be interleaved into quads, in turn, the quads could be interleaved into 8-, 12-, or 16-channel supermultiplexes. Further expansion could be made by making the submultiplexes either three or four channels rather than two channels.

To meet requirements, performance specifications include the following:

1. It should be possible to obtain up to 16 channels on the cable multiplex circuits.
2. There should be facilities for 8-channel operation on radio circuits with synchronized quads on twinplex and synchronized bichannels on polyplex.
3. On complex chains, it should be possible to use separate couplings for submultiplexes with facilities for splitting or combining submultiplexes into a supermultiplex where required.
4. All high-speed functions should be performed by electronic circuitry while submultiplex distribution and combining would be through face-plates.
5. It should be possible to adjust the channel speed of a submultiplex or supermultiplex for any rate that might be required on a standard channel between a low of 245 operations a minute or 20.4 bauds and a high of 400 or 33.9 bauds.
6. There should be no significant distortion on signals applied to transmission media.
7. There should be means of synchronizing receivers to incoming signals so that hunting will be at a minimum and it will be possible to accept reasonable gaps and periods of severe distortion without losing synchronism.
8. There should be simplicity in phasing methods and controls generally so that the equipment can be operated by ordinary attendants.
9. The synchronizing and speed-matching system should not be excessively rigid so that working can be maintained at optimum efficiency with incoming signals subject to sway and phase variation.

Description of Equipment

With the preceding criteria as a basis, the following group of equipments have been designed, produced, and placed in service:

Type *310* terminal sending; two, four, six, or eight channels.

Type *320* terminal receiving; two, four, six, or eight channels.

Type *330* electronic multiplex.

Type *340* traffic satellite for the *330* electronic multiplex.

Type *350* monitoring satellite for *310* and *330* sets.

Type 380 multiplex-to-multiplex channel couplers.

Figs. 1–16 show typical apparatus and some details.

Within the limits imposed by the use of 5- or 6-unit channels as required for specific applications, these equipments are mutually compatible with each other and all six might be in use in various places on a complex cable chain. Actually, in developing the supermultiplex and submultiplex techniques, it has been found possible to retain in service, as submultiplex receivers, old multiplex sets that have been in use for many years.

The *310* terminal sending set consists of four submultiplexes, each of two channels. The submultiplexes can be com-

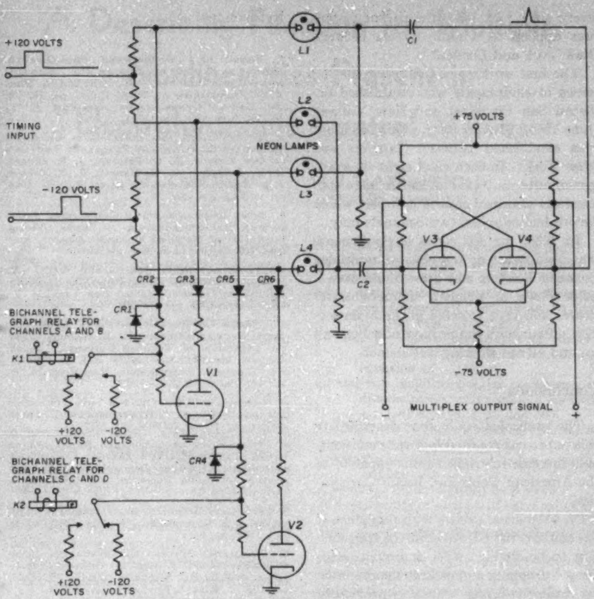


Fig. 17. Theory interleaving matrix

bined in several different ways, such as the following:

1. Four bichannels with synchronized transitions. The bichannels could be combined at an intermediate station into a supermultiplex; be applied to the four derived circuits of a polyplex radio beam; or form four separate 2-channel multiplexes.
2. Two quads with synchronized transitions. This is particularly suitable for transmitting on the two derived circuits of a twinplex radio beam; circuits of this nature have been set up between New York and Tangier and between New York and San Francisco, Calif. Separate quads are used on New York–London and London–New York cable chains.
3. One 6-channel supermultiplex of three bichannels.
4. One 8-channel supermultiplex of four bichannels.

The *320* terminal receiving is complementary to the *310* terminal sending. It can be used under the following conditions:

1. Four simultaneously keyed bichannels from a polyplex radio beam.
2. Two simultaneously keyed quads from a twinplex beam.
3. The 4-, 6-, or 8-channel supermultiplexes with suppressed singles as on a repeated cable.

The *330* electronic multiplex is a versatile set since it can be used at an intermediate station as a regenerator with facilities for dropping and picking up submultiplexes. It can also be used as a terminal station "master" transmitting or receiving set. There is no provision in this set for final distribution and printing, so satellites are required.

The *340* traffic satellite was designed to work with the *330* electronic multiplex and includes panels for receiving, transmitting, and translating from multiplex to start-stop.

The *350* monitor satellite may be connected to either type *310* or *330* sets for general monitoring.

The *380* multiplex-to-multiplex channel coupler accepts 5- or 6-unit signals from a receiving multiplex for retransmission on a sending one. Because of small variations in speed it is necessary to run the sending multiplex at a slightly higher speed than the speed of the signals being received. Blank characters are inserted to fill in the gaps. For example, the signals may be sent at 250 operations per minute which includes 247 received characters and an average of three inserted blank characters per minute.

Electronic Switching

To avoid operation of telegraph relays at a speed so high that their performance is marginal, electronic switching is employed for these high speeds. Fig. 17 is a schematic diagram illustrating one typical example of electronic switching. The telegraph relays *K1* and *K2* each operate at a speed such as to deliver the signals of one bichannel, which might be 31 bauds. The two bichannels are interleaved in the diode matrix and control the vacuum-tube fllip-flop circuit which operates at 62 bauds. These signals might be interleaved with those from another quad to form an 8-channel supermultiplex operating at 124 bauds and so on.

Electronic switching in these sets is done on a matrix basis. Some matrices are complex, others relatively simple. A typical example of one of the simpler matrices is shown in Fig. 17. This is used for interleaving two bichannels in the *310* sending set. When the positive timing pulse is applied through current-limiting resistors to diodes *CR2* and *CR3*, no current will flow through *CR2* because its cathode is at a more positive potential but current will flow through *CR3* and triode *V1* to ground because *V1* is made conducting by the positive potential applied to its grid. Therefore, neon lamp *L1* will light but *L2* will not. When lamp *L1* lights, a positive pulse is applied through capacitor *C1* to the grid of triode *V4* and causes it to conduct. It will remain conducting. When the other timing pulse reaches diode *CR5*, current flows because diode *CR4* is already made conducting by −120 volts applied via relay *K2* and current-limiting resistor to its cathode. No current will flow through diode *CR6* because triode *V2* is biased beyond cutoff. Therefore, neon lamp *L4* will light but *L3* will not. When *L4* lights, a positive pulse is applied through capacitor *C2* to the grid of triode *V3* and causes it to conduct. The multiplex output signal is taken from the anodes of *V3* and *V4*.

The principles of electronic switching as described have been employed in many ways, often in quite complex circuits.

Synchronism

The heart of any multiplex system is the means of maintaining the receiving equipment in synchronism with optimum phase of the pickup point. In the new multiplexes, a vacuum-tube-driven fork is used with frequency adjustable by means of weights and end-of-tine coils; the weights are used for coarse adjustments and the magnetizing current in the end-of-tine coils as a vernier and for phase correction. With different lengths of tines, fork frequency is variable from a low of about 30 cycles to a high of about 200 cycles per second.

Synchronism is accomplished by means of a small matrix with the "reading" pulse derived from signal transitions; routing for "fast" or "slow" is controlled by gates derived from the fork. The correction pulses are fed into an integrating capacitor which is connected to the grid of the correcting tube; this, in turn, controls the strength of the correction current which flows through fork end-of-tine coils.

This method of synchronizing has proved to be very flexible in operation since correction can be fine or coarse according to the width of the correction read pulse, value of integrating capacitor, strength of correction, etc. For chain operation it has been found preferable to have the first station of a chain use very light correction; the next station, a somewhat greater range, and so on. For point-to-point radio operation, it has not yet been shown that the potency of the phase-correction system is an important factor.

Applications

The parent system, from which the vacuum-tube fork, electronic regenerator, and synchronizing system have been taken, provided three channels in each direction between New York and London over a cable chain; this was cut into service in 1950 and has worked very efficiently every since. In 1951, the parent system was applied to a radio circuit between New York and London.

The first working circuit using the *300* series of equipments was established between San Francisco and New York in June 1954. Shortly after, a similar circuit was established between Tangiers and New York. In each case, eight 60-word-per-minute multiplex channels were successfully obtained from a twinplex which before had supplied two channels only.

In 1955, the *300* series of equipments was used to set up a 4-channel multiplex between London and New York over a cable chain in which a submersible repeater had been inserted in the main section. Four such chains have now been set up and all are working well.

Conclusions

The multiplex equipment described in this paper has resulted in a very considerable increase in trunk circuit capacity of the American Cable and Radio Corporation.

On submarine cables, with repeaters, it has enabled the effective rate of transmission to be doubled with improved facilities for dropping and picking up channels. On radio twinplex circuits, previously operated start-stop, it has quadrupled the number of available channels, without any reduction in channel speed, and without any appreciable increase in bandwidth. On both cable and radio circuits, the new equipment has resulted in appreciable improvement in working stability and the techniques developed will enable further advances to be made.

References

1. Four- or Two-Channel Time-Division-Multiplex Telegraph System for Long-Distance Radio Circuits, W. C. Peterman, A. Minc. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 362–72; also *Electrical Communication*, New York, N. Y. vol. 28, June 1951, pp. 127–41.

2. Observations and Comparisons on Radio Telegraph Signalling by Frequency Shift and On-Off Keying, H. O. Peterson, J. B. Atwood, H. E. Goldstine, G. E. Hansell, R. R. Schock. *RCA Review*, Princeton, N. J., vol. 7, Mar. 1946, pp. 11–31.

3. Some Developments in Commercial Point-to-Point Radiotelegraphy, J. A. Smale. *Journal*, Institution of Electrical Engineers, London, England, vol. 94, pt. III A, 1947, pp. 346–67.

4. Frequency Shift Telegraphy, Radio and Wire Applications, J. R. Davey, A. L. Matte. *Bell System Technical Journal*, New York, N. Y., vol. 27, Apr. 1948, pp. 265–304.

5. Design Considerations for a Radiotelegraph Receiving System, W. D. Holland. *Journal*, Institution of Electrical Engineers, London, England, vol. 98, pt. III, July 1951, pp. 253–62; also, *Electrical Communication*, New York, N. Y., vol. 29, Mar. 1952, pp. 34–50.

6. Submerged Repeaters for Long Submarine Telegraph Cables, C. H. Cramer. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1517–23.

7. The Impact of Submerged Repeaters on Global Telegraphy, C. S. Lawton. *Transactions*, Professional Group on Communication Systems, Institute of Radio Engineers, vol. CS-2, no. 3, Nov. 1954, pp. 101–12.

8. The Transmission of Intelligence in Typescript, I. S. Coggeshall. *Ibid.*, vol. CS-1, no. 1, July 1951, pp. 4–13.

9. Signalling Systems for Submarine Telegraph Circuits, C. J. Hughes. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 102, pt. B, no. 6, Nov. 1955, pp. 831–55.

10. Some Modern Techniques in Ocean Cable Telegraphy, C. H. Cramer. *AIEE Transactions*, vol. 66, 1947, pp. 494–500.

11. Twinplex and Twinmode Radio Telegraph Systems, Christopher Buff. *Electrical Communication*, New York, N. Y., vol. 29, Mar. 1952, pp. 20–33.

12. Developments in the Field of Cable and Radiotelegraph Communications, Haraden Pratt, John K. Roosevelt. *AIEE Transactions*, vol. 63, 1944, pp. 1044–47; also *Electrical Communication*, vol. 22, 1944, pp. 147–53.

◆

Discussion

John B. Moore (RCA Communications Inc., New York, N. Y.): In regard to the terminology used in the paper, the words "bichannel" or "double" are essentially adjectives rather than nouns. Other organizations, such as RCA Communications, have been using the term "diplex." "Triple" is also an adjective. I would like to suggest, for industry-wide use, and in line with the 2-channel case, the term "triplex."

The term "quad" normally is used, in the communications art, to describe other things such as two "quadded" pairs in a cable. "Quadruple" is an adjective. For consistency with 2- and 3-channel cases, and to avoid conflict with the standard telegraph term "quadruplex," I would suggest the term "quadriplex."

B. S. Diamond: Mr. Moore's comments regarding multiplex terminology have been carefully studied and discussed with various American Cable and Radio Corporation and Western Union engineers.

The terms "double," "triple," and "quad" for 2-, 3-, and 4-channel multiplexes, respectively, have been in use for many years by the Western Union, Postal Telegraph, Canadian Pacific, Canadian National, and Commercial Cable Company. These terms are straightforward and free from ambiguity. Each is listed in "Webster" as a noun. The term "diplex," is listed in "Webster" only as an adjective.

The use of "diplex" to denote a 2-channel multiplex appears objectionable because it is a word with no precise meaning and because of its association with various forms of nonmultiplex operation.

Two dissimilar applications of "diplex" may be found in Maver's "American Telegraphy,"[1] and in Herbert's "Telegraphy,"[2] respectively. The word has also been used in other applications, particularly by the Teletype Corporation in connection with a channel splitting arrangement for start-stop circuits.

The objections against the use of "diplex" do not apply against "triplex" which, actually, is a good word, synonymous with "triple." However, there is no gain in changing between synonyms, especially when "triple" is so thoroughly rooted.

"Quadriplex" does not appear in Webster and offers no advantage over "quadruple." Even if adopted, it is quite certain that all multiplex engineers and technicians will keep right on using the term "quad" for a 4-channel system.

References

1. American Telegraphy, Maver. 1892, p. 265.

2. Telegraphy, Herbert. 1930, p. 329.

A Describing Function for Multiple Nonlinearities Present in Electrohydraulic Control Valves

J. ZABORSZKY
MEMBER AIEE

H. J. HARRINGTON
ASSOCIATE MEMBER AIEE

A SPECIFIC and a general objective are jointly pursued by this paper. The specific aim is the development of a describing function[1] to represent the principal nonlinearities occurring in single-stage 4-port electrohydraulic control valves of the type sketched in Fig. 1. For maximum usefulness of the describing function towards this specific objective, it is expressed in a generalized and dimensionless form. Consequently it is applicable to the analysis or design of electrohydraulic control valves for any individual application. Precalculated curves reduce to very reasonable amounts the labor of computing the describing function for a given valve.

The general aim, or rather the achievement of general interest, is some extension of the describing function concept in two directions: 1. the simultaneous inclusion of multiple nonlinearities through an adaptation of the method of unknown coefficients, and 2. the handling of factors which have an influence on the describing function but which originate outside the circuit or mechanism for which the describing function applies. Such an influence is exerted on the describing function of the electrohydraulic valve by the phase angle and the gain of the load. The general aim of the paper is attained by illustrating the proposed methods on the special case of the hydraulic valve.

Very satisfactory agreement with analog computer studies and with a limited supply of available measured data from a physical control system is demonstrated.

Nomenclature

Any consistent system of units may be used. Primes denote derivatives with respect to τ.

A = cross-sectional area of actuator piston
A_0 = cross-sectional area of damping orifice
A_v = cross-sectional area of valve spindle
$a, b, \ldots h$ = describing function coefficients; see Appendix II, equations 10 and 11, and Figs. 2 through 9
a_n and b_n = Fourier coefficients of dimensionless valve-spindle motion, x_v; see equation 42
$\underline{a}_n$ and $\underline{b}_n$ = modified Fourier coefficients of dimensionless valve-spindle motion, x_v; (see equations 55 and 56)
$B = 2c_q K_a \cos 69$ degrees, reaction (Bernoulli) force coefficient[2,3]
c_q = discharge coefficient of valve metering orifices
e_i = voltage input to closed-loop servo
F_c = coulomb friction force of valve spindle
$F_d = \dfrac{A_v^3}{2K_v^2 A_0^2}$, orifice-damping coefficient
f_m = viscous-damping coefficient of valve spindle and torque motor
$G(ju)$ = linear transfer function of load; see equation 16
$G_L(ju)$ = linear transfer function of load; see equation 15
$G_1(ju)$ = transfer function of linearized control valve; see equation 14
$G_t(ju)$ = transfer function of torque motor electric circuit; see equation 25
I = current in torque motor
I_h = half-width of torque-motor-current hysteresis loop
$i = \dfrac{IK_m}{M_m X_{v0} \omega_n^2}$, dimensionless form of I
i_1 = dimensionless current associated with linearized control valve
$j = \sqrt{-1}$
K = adjustable amplifier gain
K_a = peripheral width of orifice in valve
$K_f = K_t K_1 K_L$, fixed part of open loop gain
K_1 = zero-frequency gain of linearized control valve
K_L = zero-frequency gain of load transfer function in equation 15
K_m = force constant of torque motor
$K_0 = KK_f$ is the open-loop gain
K_t = zero-frequency gain of torque motor electric circuit and amplifier
$K_v = \sqrt{\dfrac{1}{\rho}}$
k_m = spring constant of torque motor referred to valve spindle
M_m = mass of valve spindle plus equivalent mass of torque motor armature
$N(\psi, u)$ = describing function of electrohydraulic valve
P = accumulator-sump pressure differential
P_a = pressure differential (pressure in right chamber minus that in left chamber, Fig. 1) across actuator piston
$p = \dfrac{P_a}{P}$, dimensionless form of P_a
$q = \dfrac{\omega_n A}{T\sqrt{P}} \dfrac{X_{00}}{X_{v0}}$, dimensionless flow coefficient
$T = C_q K_v K_a$
t = time
$u = \dfrac{\omega}{\omega_n}$, dimensionless frequency
$v = \dfrac{BP}{k_m + BP}$, dimensionless reaction-force-spring nonlinearity ratio
X_a = quantity in load monitored for feedback purposes (for example, true actuator-piston motion)
X_v = valve spindle displacement; positive to right; see Fig. 1
X_{v0} = maximum valve spindle displacement from center
X_0 = idealized actuator-piston displacement for incompressible oil and no oil-ring backlash; positive displacement from center to right; see Fig. 1
X_{00} = maximum actuator-piston displacement from center
$x_a = \dfrac{X_a}{X_{00}}$, dimensionless form of X_a
$x_v = \dfrac{X_v}{X_{v0}}$, dimensionless form of X_{v0}
$x_0 = \dfrac{X_0}{X_{00}}$, dimensionless form of X_0
ϵ = servo error
$\zeta = \dfrac{f_m}{2\sqrt{M_m(k_m + BP)}}$, dimensionless form of f_m
$\zeta_c = \dfrac{F_c}{2M_m X_{v0} \omega_n^2}$, dimensionless form of F_c
$\zeta_d = \dfrac{F_d X_{v0}}{2M_m}$, dimensionless form of F_d
θ = phase angle of $G(ju)$
ξ = dimensionless amplitude of $G(ju)$
ρ = oil density
$\tau = \omega_n t$, dimensionless time
ϕ = phase angle of base harmonic of x_v
ψ = peak amplitude of sinusoidal oscillations in pressure ratio p
ω = circular frequency of a sinusoidal oscillation
$\omega_n = \sqrt{\dfrac{k_m + BP}{M_m}}$ is the undamped natural frequency of the linearized valve

Nonlinearities Included

The describing function proposed here is believed to be much more complete than any published heretofore for electrohydraulic control valves, since it simultaneously includes all principal nonlinearities of such valves. Specifically, these include the nonlinear flow versus pressure drop relationship at the valve orifices; the reaction (Bernoulli) forces acting on the valve spindle as a result of the flow through the valve orifices; orifice-type damping of the spindle, if any (see Fig. 1); and coulomb friction of the valve spindle. Torque-motor hysteresis or backlash as such is not included but analog computer studies[4] indicate that the

---

Paper 57-107, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted July 24, 1956; made available for printing December 4, 1956.

J. ZABORSZKY, is with Washington University, St. Louis, Mo., and is consultant to McDonnell Aircraft Corporation, St. Louis, Mo.; H. J. HARRINGTON is with McDonnell Aircraft Corporation, St. Louis, Mo.

effect of this nonlinearity is very similar to the effect of the coulomb friction acting on the valve spool (at least as long as the backlash has an approximately straight-line rhomboidal characteristic), and so it can be equivalently replaced by coulomb friction in applying the method of this paper. The basis of equivalent substitution is that the effective coulomb friction coefficient F_c be made equal to the torque-motor force coefficient K_m times the half-width of the current hysteresis loop I_h, i.e.,

$$F_c = K_m I_h \tag{1}$$

Nondimensional Parameters

Nondimensional parameters directly involved in the computation of the describing function are defined in the following, and can be computed from the actual valve dimensions and system parameters by these equations:

$$\omega_n = \sqrt{\frac{k_m + BP}{M_m}} \tag{2}$$

$$u = \frac{\omega}{\omega_n} \tag{3}$$

$$v = \frac{BP}{k_m + BP} \tag{4}$$

$$\zeta = \frac{f_m}{2\sqrt{M_m(k_m + BP)}} \tag{5}$$

$$\zeta_c = \frac{F_c}{2M_m X_{e0} \omega_n^2} \tag{6}$$

$$\zeta_d = \frac{F_d X_{e0}}{2M_m} \tag{7}$$

$$q = \frac{\omega_n A}{T} \sqrt{\frac{1}{P}} \frac{X_{00}}{X_{e0}} \tag{8}$$

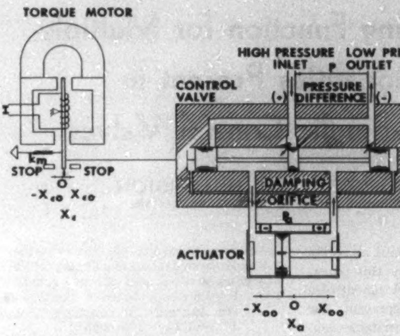


Fig. 1. Sketch of a single-stage 4-port electrohydraulic valve and torque motor

Equations and Precalculated Curves for Computing the Describing Function

The describing function is derived in the Appendixes and, in dimensionless form, is given by

$$N(\psi, u) = \sqrt{\frac{(1-u^2)^2 + (2\zeta u)^2}{C^2 + S^2}} \times \underline{\Big/ \theta + \tan^{-1}\frac{S}{C} + \tan^{-1}\frac{2\zeta u}{1-u^2} + 90 \text{ degrees}} \tag{9}$$

where

$$C = a(u^2-1) - b2\zeta u + cv\psi - dq\zeta_d u^3 \xi\psi - e\frac{2\zeta_c}{qu\xi\psi} \tag{10}$$

$$S = b(u^2-1) + a2\zeta u + fv\psi + gq\zeta_d u^3 \xi\psi + h\frac{2\zeta_c}{qu\xi\psi} \tag{11}$$

The coefficients *a*, *b*, *c*, *d*, *e*, *f*, *g*, and *h* are defined in Appendix II, and are plotted as functions of ψ and θ in Figs. 2 through 9. The parameter ψ is the peak value of the dimensionless actuator pressure differential p when it is varying sinusoidally, and so ψ establishes the amplitude of possible oscillations. The angle θ is the phase and ξ is the amplitude of the dimensionless linear transfer function, $G(ju)$, which links the idealized (incompressible oil and no oil-ring backlash) actuator piston motion x_0 as output to the pressure differential p across the actuator as input; see equation 16. This transfer function is determined from the compressibility of the oil and the other dynamic properties of the load by entirely conventional methods. (To avoid an uncertainty of ±180 degrees in the value of θ, the transfer function of the load must be based on the reference directions which are selected for x_0 and p in the Nomenclature. With these reference directions, θ will be a positive angle for all practical loads.) The rest of the coefficients in equations 9 through 11 represent the data of the valve as defined in equations 2 through 8.

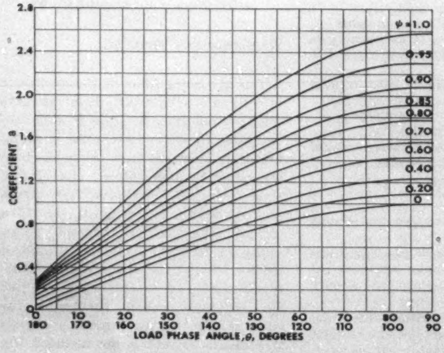


Fig. 2. Describing function coefficient a

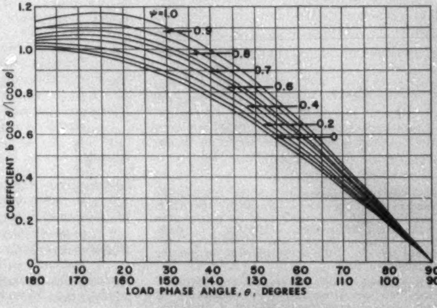


Fig. 3. Describing function coefficient b

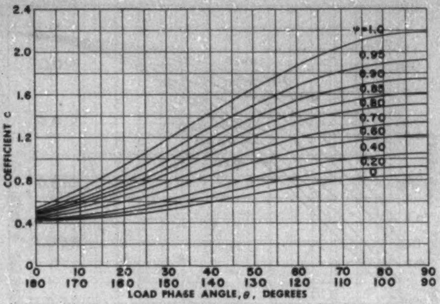


Fig. 4. Describing function coefficient c

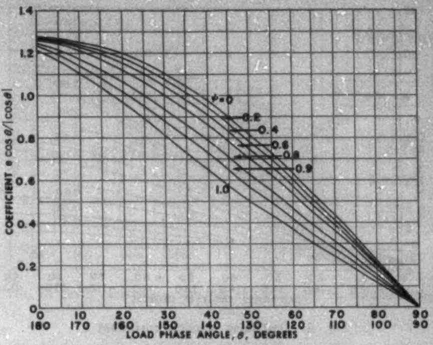


Fig. 6. Describing function coefficient e

Interpretation of the Describing Function

The describing function is inserted into the equivalent system diagram as is shown in Fig. 10, ahead of the linearized control valve.

The linearized control valve is defined as one having the same physical dimensions as the actual control valve but having a flow rate which is simply proportional to the valve opening

$$A\dot{X}_0 = T\sqrt{P}X_e \tag{12}$$

and a spring constant of

$$k_m + BP \tag{13}$$

but no nonlinear reaction (Bernoulli) forces, no coulomb friction or orifice damping, or any other nonlinearities. The dimensionless and normalized transfer function representing such a linearized valve is, of course,

$$K_1G_1(ju) = \frac{x_0}{i_1} = \frac{1/q}{(ju)[(ju)^2 + 2\zeta ju + 1)]} \tag{14A}$$

$$K_1 = 1/q \tag{14B}$$

Two transfer functions are needed for the load. The first gives the quantity monitored for feedback purposes x_a (true actuator-piston motion, for example) as output against idealized actuator-piston motion x_0 as input:

$$K_LG_L(ju) = \frac{x_a}{x_0} \tag{15}$$

This transfer function is to be applied in series with the describing function in the manner shown by the block diagram of Fig. 10. The other transfer function, which must be dimensionless and is required for the computation of the describing function (equations 9 through 11), relates the idealized actuator piston motion x_0 to the pressure drop across the piston p:

$$G(ju) = \frac{x_0}{p} = \xi\underline{/\theta} \tag{16}$$

Both of these transfer functions should include the complete dynamic system which is controlled by the valve including the compressibility of the oil and any mass, damping, and linkage-elasticity effects of the actuator and all elements driven directly or indirectly by the actuator. Any internal feedbacks existing within the load proper must be included. However, it should be remembered that both $G_L(ju)$ and $G(ju)$ are open-loop transfer functions in the sense that they include no feedbacks which close directly or indirectly through the valve.

The complexity of the load or the order of the transfer functions $G_L(ju)$ and $G(ju)$ is not restricted. The only restriction on the load is that it can be represented by transfer functions which are established by standard linear or linearization techniques. Use of the describing function for calculating amplitudes and frequencies of sustained oscillations does not differ from usual practices.

Illustrative Example

Consider a hydraulic valve the nondimensional constants of which are

$$q = 68.5 \tag{17}$$

$$\zeta = 0.0441 \tag{18}$$

$$\zeta_c = 0.0159 \tag{19}$$

$$\zeta_d = 6.30 \tag{20}$$

$$v = 0.25 \tag{21}$$

so that, in agreement with equation 14, it

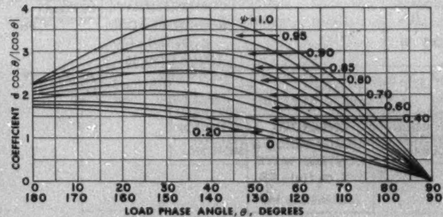


Fig. 5. Describing function coefficient d

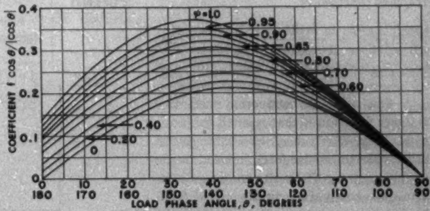


Fig. 7. Describing function coefficient f

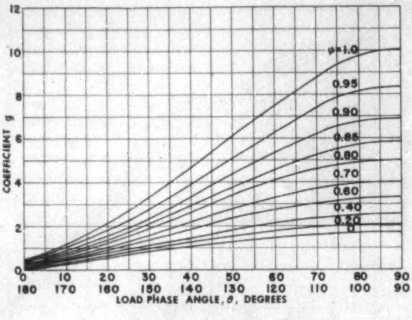


Fig. 8. Describing function coefficient g

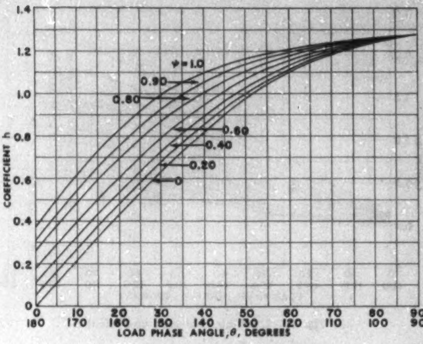


Fig. 9. Describing function coefficient h

has a dimensionless linearized transfer function of

$$K_1G_1(ju)=\frac{x_0}{i_1}=\frac{0.0146}{ju[(1-u^2)+0.0882ju]} \tag{22A}$$

$$K_1=0.0146 \tag{22B}$$

The valve is to be used in connection with a load having a nondimensional transfer function, equation 16, of

$$G(ju)=\frac{x_0}{p}=\xi\underline{/\theta}=0.603\times\frac{(-1+20.8u^2-0.464u^4)-j(0.423u-1.55u^3)}{(1-271.5u^2+26.4u^4)+j(9.64u-88.6u^3)} \tag{23}$$

The frequency response for this transfer function is shown in Fig. 11. As is shown by the system block diagram of Fig. 12, and by equation 15, another transfer function of the load is needed to link x_a to x_0, namely

$$K_LG_L(iu)=\frac{x_a}{x_0}=(0.982)\times\frac{(1-16.3u^2)+0.258ju}{(1-20.8u^2+0.464u^4)+j(0.423u-1.55u^3)} \tag{24A}$$

$$K_L=0.982 \tag{24B}$$

The electric circuit (inductance and resistance) of the torque motor will have a transfer function of

$$K_tG_t(ju)=\frac{i}{\epsilon}=\frac{1.19}{0.468ju+1} \tag{25A}$$

$$K_t=1.19 \tag{25B}$$

A fine point in connection with this transfer function is that a small feedback exists, Fig. 12, from x_ϵ' to i through the counterelectromotive force induced in the torque motor coil by the motion of the armature and valve spindle. This feedback produces an inner loop around the block of the describing function. The inner loop could be handled by conventional describing function techniques but to keep the present discussion brief it will be neglected. The main loop is closed through a straight unity feedback; however, a network in the feedback could be included with equal ease.

One point of the describing function will be calculated here for illustration, namely, that corresponding to pressure ratio

$$\psi=0.8 \tag{26}$$

and frequency

$$u=0.210 \tag{27}$$

From Fig. 11 it is easily found that, at $u=0.210$

$$\xi=0.00620 \text{ and } \theta=48.0 \text{ degrees} \tag{28}$$

Then, from Figs. 2 through 9, for $\psi = 0.8$ and $\theta = 48.0$ degrees, there is found

$$a=1.34;\ b=0.810;\ c=1.18 \tag{29}$$

$$d=2.38;\ e=0.660;\ f=0.295 \tag{30}$$

$$g=3.30;\ h=1.085 \tag{31}$$

Substituting equations 17 through 21, and equations 26 through 31 into equations 10 and 11,

$$C=-1.404 \tag{32}$$

$$S=-0.140 \tag{33}$$

Then, with equation 9, the describing function is

$$N(0.8,\ 0.210)=0.68\underline{/-35.2 \text{ degrees}} \tag{34}$$

or

$$-\frac{1}{N(0.8,\ 0.210)}=3.36\ db\underline{/-144.8 \text{ degrees}} \tag{35}$$

Fig. 13 shows $-1/N\ (\psi,\ u)$ loci for several values of ψ and u for the illustrative example. The value just calculated is shown by a darkened triangle on the $u=0.210$ locus at $\psi = 0.8$. Other points of these loci were calculated in the same manner as the one point illustrated.

In accordance with usual describing-function[1] and Nichols chart techniques, the $KK_fG_t(ju)G_1(ju)G_L(ju)$ locus is best plotted on a diaphanous sheet on the same

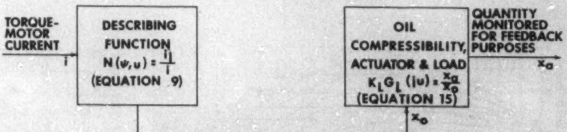
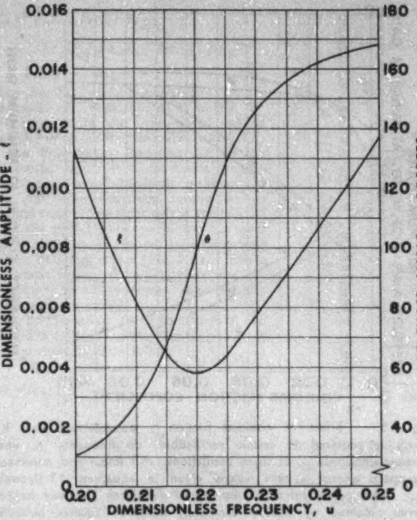

TORQUE-MOTOR CURRENT
i
DESCRIBING FUNCTION
$N(\psi,u)=\frac{i_1}{i}$
(EQUATION 9)
IDEALIZED TORQUE-MOTOR CURRENT
i_1
LINEARIZED TORQUE-MOTOR & CONTROL VALVE
$K_1G_1(ju)=\frac{x_0}{i_1}$
(EQUATION 14)
IDEALIZED ACTUATOR PISTON MOTION
x_0
OIL COMPRESSIBILITY, ACTUATOR & LOAD
$K_LG_L(ju)=\frac{x_a}{x_0}$
(EQUATION 15)
QUANTITY MONITORED FOR FEEDBACK PURPOSES
x_a

Fig. 10. Block diagram showing the way the describing function is to be inserted



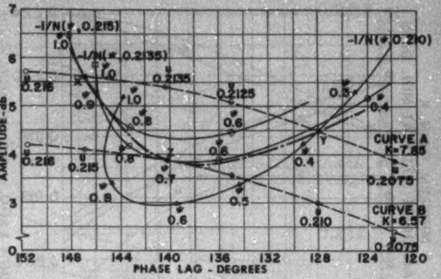


Fig. 11 (left). Frequency response of $G(ju)=\xi/\underline{\theta}$ for the illustrative example

Fig. 13 (above). Sample application of the describing function in Nichols chart of illustrative example

scale as that for the describing function; see Fig. 13. By superimposing this diaphanous sheet on the describing function with ordinate axes lined up, the effect of variations in the amplifier gain can be considered by vertically shifting the sheets so that every relative position of the $-1/N(\psi,u)$ loci and the $KK_fG_f(ju)G_1(ju)$ $G_L(ju)$ locus corresponds to one gain value. In Fig. 13, the $KK_fG_f(ju)G_1(ju)G_L(ju)$ locus is shown for two amplifier-gain values. Curve *A*, which represents a gain of $K = 7.85$, shows two like-frequency intersections with the $-1/N(\psi, u)$ loci at the frequencies $u = 0.215$ and $u = 0.210$ (points *X* and *Y* respectively). Such intersections indicate the possibility of sustained, or limit-cycle, oscillations.[1] Furthermore, the intersection at $u = 0.215$, point *X*, provides a feasible sustained oscillation, while that at $u = 0.210$, point *Y*, an unfeasible one. The sustained oscillation at point *X* is feasible since a reduction of amplitude ψ carries the oscillation below the curve *A* indicating system instability, and so the resulting increasing amplitude returns the system into its oscillation at point *X*. Conversely, an increase of amplitude ψ from its value at *X* shifts the system into the stable region so that the increase is damped out immediately. Consequently, point *X* indicates a condition where the system oscillation can be indefinitely sustained. On the other hand, sustained oscillation at point *Y* is not practically feasible since the crossing of the two curves is reversed at *Y* relative to the crossing at *X*. Thus, a repetition of the foregoing discussion shows that, at the point *Y*, either a decrease or an increase of the amplitude, however slight initially, becomes cumulative.

Now, the behavior of the system with amplifier gain $K=7.85$ can be described on the basis of these results of Fig. 13. At this gain the system will not spontaneously start to oscillate; in fact oscillations started by outside disturbances will die out unless their amplitude exceeds $\psi=0.37$ (point *Y*). If, however, the system is excited to an oscillation of amplitude larger than $\psi=0.37$, this oscillation will either expand or attenuate until its amplitude becomes $\psi=0.93$ (point *X*), where it will be sustained. With changing gain, points *X* and *Y* shift along the dash-dot curves of Fig. 13 (as may be ascertained by plotting intermediate loci of both sets). When the gain is reduced, points *X* and *Y* approach each other and finally, at gain $K=6.57$ (curve *B*), they merge at point *Z*. Further reduction of gain results in no more meaningful intersections between the two sets of loci. So, $K = 6.57$ is the minimum amplifier gain for the given system where sustained oscillation is marginally feasible; below this gain the system is stable in the normal linear sense. This value of gain is shown by a solid dot in Fig. 14. The accompanying amplitude $\psi=0.71$ and the frequency $u=0.2135$ from point *Z* in Fig. 13 are also shown by solid dots in Fig. 14. Other points of the calculated curves in Figs. 14 through 16 were similarly obtained.

Effect of Certain Valve Parameters

A study of the effects of certain linear and nonlinear parameters, such as the Bernoulli force, coulomb friction, and dashpot damping, gives important information for the designer and also provides an opportunity to test the accuracy of the proposed describing function against analog computer data and data obtained from an actual system. Such a study is summarized in Figs. 14 through 16. Only one nonlinear parameter is varied in each of these curves; how-

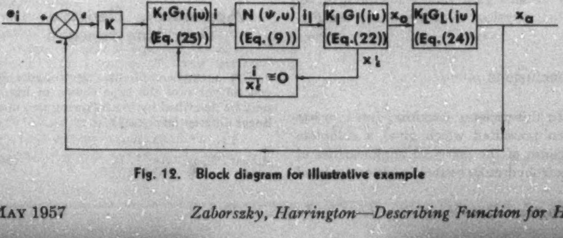


Fig. 12. Block diagram for illustrative example

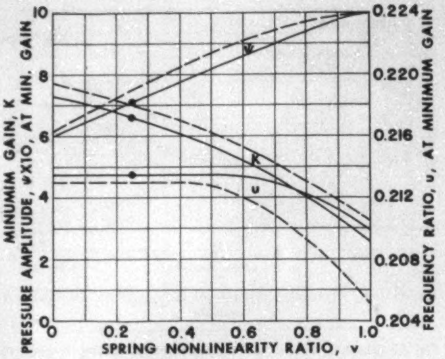


Fig. 14. **Effect of spring nonlinearity ratio v on minimum gain K which is required to sustain oscillations; on frequency u, and pressure-amplitude ψ, of these oscillations. All linear and nonlinear parameters except v have values given in equations 17 through 21. Solid curves calculated by use of describing function; dashed curves obtained by analog computer**

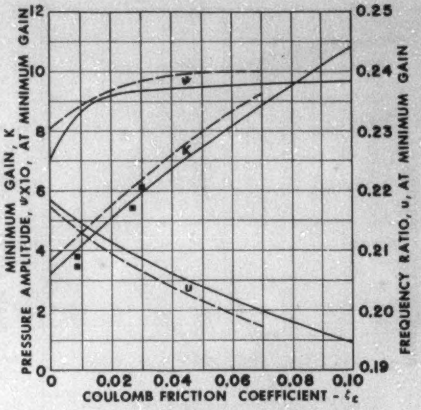


Fig. 15. **Effect of coulomb friction ζ_c on minimum gain K, which is required to sustain oscillations; on frequency u, and pressure-amplitude ψ, of these oscillations. All linear and nonlinear parameters except ζ_c have values given in equations 17 through 21. Solid curves calculated by use of describing function; dashed curves obtained by analog computer; darkened squares indicate measurements from a physical electrohydraulic servo system**

ever, all the other nonlinearities are simultaneously present and retain their constant, nominal values as given in equations 17 through 21.

The reaction or Bernoulli forces on the valve (the last two terms in equation 37) are proportional to the opening of the valve X_e and to the pressure drop across the valve $(P+P_A)$. Consequently, at a constant pressure drop these forces would behave as an ordinary linear spring. Since, however, the pressure drop is determined by the instantaneous phase of the motion of the load, the Bernoulli force represents a nonlinear feedback from the load to the valve, which feedback has a destabilizing effect. The existence of this destabilizing effect can be proved and its extent studied by the proposed describing function.

The effect of the Bernoulli, or reaction, force on the valve spindle will depend on its magnitude relative to the linear spring force on the valve spindle. This relative magnitude is conveniently described by the dimensionless parameter v, the spring nonlinearity ratio. Parameter v gives the fraction of the total apparent spring constant (k_m+BP) which is derived from the Bernoulli force BP when the pressure drop across the actuator is zero.

Fig. 14 indicates that an increase in the nonlinear feedback represented by the Bernoulli force results in very substantial reductions of the permissible open-loop gain and in considerable increases in the amplitude of the sustained oscillations. Thus, large values of the Bernoulli force are detrimental for stability even if they are beneficial for the steady characteristics of the valve.[5]

In a similar manner Fig. 15 shows that coulomb friction at the valve spindle has a stabilizing effect at the mode of oscillation near the resonant frequency of the load. It should be remembered however that there is another mode attributable to the coulomb friction itself.[4] This latter mode can also be studied with the proposed describing function. Finally, Fig. 16 shows the essentially stabilizing effects of another nonlinear parameter, dashpot damping.

The excellent agreement shown in Figs. 14 through 16 between analog computer solutions[4] and those supplied by the describing function for corresponding conditions is evidence of the latter's accuracy. Very satisfactory accuracy is also indicated by comparing calculated minimum gain values with the rather meager number of such gain values measured on an actual physical control system; see Fig. 15.

Conclusions

In this paper a describing function has been presented which gives a complete account of the principal nonlinearities of electrohydraulic valves. Precalculated curves reduce the labor for computing the describing function of a specific valve to a very reasonable amount. It is shown that this describing function is a powerful tool for the study of the effect of linear and nonlinear valve parameters on the dynamic operation of the valve with given loads. Such studies are helpful in selecting the best valve characteristics.

The paper also represents a generalization of the describing function method in the way it treats multiple nonlinearities, and especially in the way it includes the effect of the load which is an element apart from the valve for which the describing function is being developed.

The paper is restricted to single-stage valves and to loads which can be represented by linear transfer functions. Later publications are being planned to deal with 2-stage valves and nonlinear loads.

Appendix I. Derivation of Describing Function

The operation of an electrohydraulic control valve of the type shown in Fig. 1 may be described by the following two nonlinear differential equations:

$$A\dot{X}_0=TX_e\sqrt{P+\frac{X_e}{|X_e|}P_a} \qquad (36)$$

$$IK_m = M_m\ddot{X}_e + f_m\dot{X}_e + F_c\frac{\dot{X}_e}{|\dot{X}_e|} + F_d\dot{X}_e|\dot{X}_e| + k_mX_e + BPX_e + BP_a|X_e| \quad (37)$$

where X_0, X_e, I, and P_a, as well as time, are the variables. Equation 36 is simply the flow relation for the sharp-edged orifices resulting upon opening the valve, where $A\dot{X}_0$ is the flow rate. Equation 37 is the differential equation of the valve spindle; the left-hand term is the force on the spindle from the electromagnetic torque motor; the first term on the right side is the inertial force, the second is viscous friction, the third coulomb friction, the fourth damping caused by the damping orifice (Fig. 1), the fifth is the linear spring restraining force and the last two terms represent the reaction (Bernoulli) forces.[2,3,5]

By nondimensionalizing equations 36 and 37 with the aid of equations 2 through 8, they can be rearranged as follows:

$$qx_0' = x_e\sqrt{1+\frac{x_e}{|x_e|}p} \quad (38)$$

$$i = x_e'' + 2\zeta x_e' + 2\zeta_c\frac{x_e'}{|x_e'|} + 2\zeta_d x_e'|x_e'| + x_e + v|x_e|p \quad (39)$$

where i, x_0, x_e, and p, as well as dimensionless time, are the variables.

The describing function is defined as the complex ratio of the base harmonic of the output to the base harmonic of the input across a nonlinear component. Normally it is assumed that the input is purely sinusoidal. The low-pass character of servomechanisms then justifies neglecting higher harmonics of the output.

In the present case when the input i was assumed purely sinusoidal, solution was found to become mathematically unfeasible. For mathematical expediency (which is subsequently justified by excellent agreement with analog computer studies including all harmonics) it is assumed that

$$p = \psi\cos u\tau \quad (40)$$

is purely sinusoidal. Then, also

$$x_0 = \xi\psi\cos(u\tau+\theta) \quad (41)$$

since p and x_0 are linked only through the load which is assumed to be linear; see equation 16. However, x_e and i then must be represented by Fourier series

$$x_e = \sum_{n=1}^{\infty}(a_n\cos nu\tau + b_n\sin nu\tau) \quad (42)$$

$$i = \sum_{m=1}^{\infty}(i_{Cm}\cos mu\tau + i_{Sm}\sin mu\tau) \quad (43)$$

where the a_n, b_n, i_{Cm}, and i_{Sm} coefficients are so far unknown.

Now, if the binomial expansion is used in equation 38 (considering that $(x_e/|x_e|)^{2n}=1$ for integer n) and if equation 40 is substituted into equation 38 and certain trigonometric identities are applied, there results

$$qx_0' = x_e[A_0(\psi) - A_2(\psi)\cos 2u\tau - A_4(\psi)\cos 4u\tau - A_6(\psi)\cos 6u\tau\ldots] + |x_e|[A_1(\psi)\cos u\tau + A_3(\psi)\cos 3u\tau + A_5(\psi)\cos 5u\tau + \ldots] \quad (44)$$

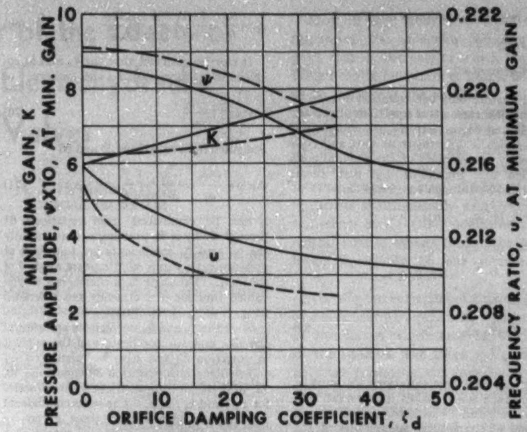


Fig. 16. Effect of orifice damping coefficient ζ_d on minimum gain K, which is required to sustain oscillations; on frequency u, and pressure-amplitude ψ, of these oscillations. All linear and nonlinear parameters, except ζ_d and q=97.0, have values given in equations 17 through 21. Solid curves calculated by use of describing function; dashed curves obtained by analog computer

where the $A_1(\psi)$ coefficients are polynominals in ψ as follows:

$$\begin{aligned}
A_0(\psi) &= 1-0.0625\psi^2-0.0145\psi^4-0.0064\psi^6-\ldots\\
A_1(\psi) &= 0.5\psi+0.047\psi^3+0.0171\psi^5+\ldots\\
A_2(\psi) &= 0.063\psi^2+0.0195\psi^4+0.0096\psi^6+\ldots\\
A_3(\psi) &= 0.0156\psi^3+0.00855\psi^5+\ldots\\
A_4(\psi) &= 0.00488\psi^4+0.00384\psi^6+\ldots\\
A_5(\psi) &= 0.00171\psi^5+\ldots\\
A_6(\psi) &= 0.00064\psi^6+\ldots
\end{aligned} \quad (45)$$

An expression is still needed for $|x_e|$. This will also consist of harmonics but these unfortunately can not be calculated in a general way from the a_n and b_n coefficients of equation 42. Since, according to equations 45, the first bracket in equation 44 is larger than the second bracket, a relatively cruder approximation for $|x_e|$ than for x_e is justified. Furthermore, computed values of a_n and b_n bear out that a rather good approximation is to determine $|x_e|$ from only the first harmonic of x_e. Therefore $|x_e|$ may be approximated as

$$|x_e| = \alpha_0 + \sum_{r=1}^{\infty}(\alpha_{2r}\cos 2ru\tau + \beta_{2r}\sin 2ru\tau) \quad (46)$$

where the coefficients α_0, α_{2r}, and β_{2r} are defined as

$$\alpha_0 = \frac{2}{\pi}\sqrt{a_1^2+b_1^2} = \frac{2}{\pi}\frac{a_1}{\cos\phi} \text{ where}$$

$$\phi = \tan^{-1}\frac{-b_1}{a_1} \quad (47)$$

$$\alpha_{2r} = \frac{(-1)^{r+1}}{(4r^2-1)}\frac{4}{\pi}\sqrt{a_1^2+b_1^2}\cos 2r\phi = \frac{(-1)^{r+1}}{(4r^2-1)}\frac{4}{\pi}\frac{a_1}{\cos\phi}\cos 2r\phi \quad (48)$$

$$\beta_{2r} = \frac{(-1)^{r}}{(4r^2-1)}\frac{4}{\pi}\sqrt{a_1^2+b_1^2}\sin 2r\phi = \frac{(-1)^{r+1}}{(4r^2-1)}\frac{4}{\pi}\frac{b_1}{\sin\phi}\sin 2r\phi \quad (49)$$

If now equations 41 and 42 as well as equations 46 through 49 are substituted into equation 44 there results

$$-q\xi\psi u\sin(u\tau+\theta) = -q\xi\psi u\sin\theta\cos u\tau - q\xi\psi u\cos\theta\sin u\tau = \sum_{s=1}^{\infty}L_s(a_1, a_2, \ldots a_n)\cos su\tau + \sum_{s=1}^{\infty}N_s(b_1, b_2, \ldots b_n)\sin su\tau \quad (50)$$

where the $L_s(a_1, a_2, \ldots a_n)$ are linear combinations of the coefficients $a_1, a_2, \ldots a_n$, while the $N_s(b_1, b_2, \ldots b_n)$ are linear combinations of the coefficients $b_1, b_2, \ldots b_n$.

But equation 50 can be satisfied for all values of τ only if

$$-q\xi\psi u\sin\theta = L_1(a_1, a_2, \ldots a_n) \quad (51)$$

$$0 = L_s(a_1, a_2, \ldots a_n) \quad (52)$$

$$s = 2, 3, 4, \ldots. n$$

$$-q\xi\psi\cos\theta = N_1(b_1, b_2, \ldots b_n) \quad (53)$$

$$0 = N_s(b_1, b_2, \ldots b_n) \quad (54)$$

$$s = 2, 3, 4, \ldots. n$$

Equations 51 and 52 represent a set of n linear algebraic equations for the coefficients $a_1, a_2, \ldots a_n$ while equations 53 and 54 form a similar set of n equations for $b_1, b_2, \ldots b_n$. (Strictly speaking $n\rightarrow\infty$ but, whatever the desired accuracy, there will always be a finite n beyond which the higher harmonics become negligible.)

If, now, the notation is introduced

$$a_n = \underline{a}_n q\xi\psi u \sin\theta \quad (55)$$

$$b_n = \underline{b}_n q\xi\psi u \cos\theta \quad (56)$$

then, because only one equation of each of the simultaneous sets of equations 51 and 52, and 53 and 54 respectively has an absolute term which is different from zero, and consequently all members of the simultaneous solution are multiplied by that term, equations 51 through 54 change into

$$-1 = L_1(\underline{a}_1, \underline{a}_2, \ldots \underline{a}_n) \quad (57)$$

$$0 = L_s(\underline{a}_1, \underline{a}_2, \ldots \underline{a}_n) \quad (58)$$

$$s = 2, 3, 4 \ldots n$$

$$-1 = N_1(\underline{b}_1, \underline{b}_2, \ldots \underline{b}_n) \quad (59)$$

$$0 = N_s(\underline{b}_1, \underline{b}_2, \ldots \underline{b}_n) \quad (60)$$

$$s = 2, 3, 4 \ldots n$$

Writing out these equations for some n, for instance $n=6$, makes it apparent that all even a's and b's are zero. Furthermore, a study of the relative magnitudes of the various terms in the equations containing the odd coefficients shows that the following approximate equations yield the odd coefficients with good accuracy:

$$\underline{a}_1 = \frac{-2}{2A_0 - A_2 + 2A_1 s_0 + (A_1+A_3)s_2 + (A_3+A_5)s_4 + A_5 s_6} \quad (61)$$

$$\underline{a}_3 = -\frac{-A_2 - A_4 + 2A_3 s_0 + (A_1+A_5)s_2 + A_1 s_4 + A_3 s_6}{2A_0 - A_6}\,\underline{a}_1 \quad (62)$$

$$\underline{a}_5 = -\frac{-A_4 - A_6 + 2A_5 s_0 + A_3 s_2 + A_1(s_4+s_6)}{2A_0}\,\underline{a}_1 \quad (63)$$

$$\underline{b}_1 = \frac{-2}{2A_0 + A_2 + (A_1-A_3)c_2 + (A_3-A_5)c_4 + A_5 c_6} \quad (64)$$

$$\underline{b}_3 = \frac{-A_2 + A_4 + (A_1-A_5)c_2 + A_1 c_4 + A_3 c_6}{2A_0 + A_6}\,\underline{b}_1 \quad (65)$$

$$\underline{b}_5 = -\frac{-A_4 + A_6 + A_3 c_2 + A_1 c_4 + A_1 c_6}{2A_0}\,\underline{b}_1 \quad (66)$$

where the A_i terms are those given by equations 49 and the s_0, s_{2r} and c_{2r} terms are defined as

$$s_0 = \frac{\alpha_0}{a_1} = \frac{2}{\pi}\frac{1}{\cos\phi} \quad (67)$$

$$s_{2r} = \frac{\alpha_{2r}}{a_1} = \frac{(-1)^{r+1}4}{(4r^2-1)\pi}\frac{\cos 2r\phi}{\cos\phi} \quad (68)$$

$$c_{2r} = \frac{\beta_{2r}}{b_1} = \frac{(-1)^{r+1}4}{(4r^2-1)\pi}\frac{\sin 2r\phi}{\sin\phi} \quad (69)$$

It can be observed that the $\underline{a}_n$ and $\underline{b}_n$ coefficients are functions of ψ and ϕ where

$$\phi = \tan^{-1}\frac{-b_1}{a_1} \quad (70)$$

But since from equations 55 and 56

$$\tan\phi = \frac{-b_1}{a_1} = \frac{-\underline{b}_1}{\underline{a}_1}\frac{1}{\tan\theta} \quad (71)$$

ϕ can be eliminated from equations 61 through 66 by cross plotting so that actually the $\underline{a}_n$ and $\underline{b}_n$ coefficients are functions of pressure ratio ψ and of the phase angle θ of the load. The angle θ, of course, is a single-valued function of frequency and is known for a given load; see equation 16.

Once the $\underline{a}_n$ and $\underline{b}_n$ coefficients are known, the i_{Cm} and i_{Sm} coefficients of the current in equation 43 can also be calculated by substituting into equation 39 equations 40, 42, and 44 and appropriate Fourier series for $|x_e'|$ and $x_e'/|x_e'|$. The series coefficients for $|x_e'|$ are identical to those given by equations 45 through 47 for $|x_e|$ except that each term is multiplied by u and the $r=1, 3, 5\ldots$ terms change sign. The Fourier series used to approximate $x_e'/|x_e'|$ is simply that of a square wave of unit amplitude which is in phase with the base harmonic of x_e'. Thus

$$\frac{x_e'}{|x_e'|} = -\frac{4}{\pi}\Big[\cos\phi\sin u\tau + \frac{1}{3}\cos 3\phi\sin 3u\tau + \frac{1}{5}\cos 5\phi\sin 5u\tau\ldots + \sin\phi\cos u\tau + \frac{1}{3}\sin 3\phi\cos 3u\tau + \frac{1}{5}\times \sin 5\phi\cos 5u\tau + \ldots\Big] \quad (72)$$

For the first harmonic of i

$$i_{C_1} = q\xi\psi u C \quad (73)$$

$$i_{S_1} = q\xi\psi u S \quad (74)$$

where C, S, and the other notations are defined in equations 10, 11, and in Appendix II. So, retaining the first harmonic only

$$i = q\xi\psi u(C\cos u\tau + S\sin u\tau) \quad (75)$$

or in complex form

$$\bar{i} = q\xi\psi u\sqrt{C^2+S^2}\;\angle -\tan^{-1}\frac{S}{C} \quad (76)$$

Now, if the valve were linear in the sense of the "linearized valve" of Fig. 10 and equation 14, it would be described by the following linear differential equations rather than in equations 38 and 39:

$$i_1 = x_e'' + 2\zeta x_e' + x_e \quad (77)$$

$$qx_e' = x_e \quad (78)$$

Equation 41 still holds for the load. This can be solved as

$$\bar{i}_1 = q\xi\psi u\sqrt{(1-u^2)+(2\zeta u)^2}\;\angle\,\theta + 90 \text{ degrees} + \tan^{-1}\frac{2\zeta u}{1-u^2} \quad (79)$$

Dividing equation 79 by equation 76 results in the describing function defined in equation 9 and Fig. 10.

Appendix II

Definitions of coefficients used in the describing function equation are as follows:

$$a = -\underline{a}_1\sin\theta \quad (80)$$

$$b = -\underline{b}_1\cos\theta \quad (81)$$

$$c = \left(s_0 + \frac{1}{2}s_2\right)\underline{a}_1\sin\theta = -\left(s_0 + \frac{1}{2}s_2\right)a \quad (82)$$

$$d = -[2s_0\underline{b}_1 - s_2(\underline{b}_1 + 3\underline{b}_3) + s_4(3\underline{b}_3 + 5\underline{b}_5) - s_6 5\underline{b}_5 + \ldots]\underline{a}_1\sin\theta\cos\theta - [c_2(\underline{a}_1 + 3\underline{a}_3) - c_4(3\underline{a}_3 + 5\underline{a}_5) + c_6 5\underline{a}_5 + \ldots]\underline{b}_1\sin\theta\cos\theta \quad (83)$$

$$e = \frac{4}{\pi}\sin\left[\tan^{-1}\left(\frac{-\underline{b}_1}{\underline{a}_1}\frac{1}{\tan\theta}\right)\right] \quad (84)$$

$$f = \frac{1}{2}c_2\underline{b}_1\cos\theta = -\frac{1}{2}c_2 b \quad (85)$$

$$g = -[2s_0\underline{a}_1 + s_2(\underline{a}_1 - 3\underline{a}_3) - s_4(3\underline{a}_3 - 5\underline{a}_5) + s_6 5\underline{a}_5 + \ldots]\underline{a}_1\sin^2\theta - [c_2(\underline{b}_1 - 3\underline{b}_3) - c_4(3\underline{b}_3 - 5\underline{b}_5) + c_6 5\underline{b}_5 + \ldots]\underline{b}_1\cos^2\theta \quad (86)$$

$$h = -\frac{4}{\pi}\cos\left[\tan^{-1}\left(\frac{-\underline{b}_1}{\underline{a}_1}\frac{1}{\tan\theta}\right)\right] \quad (87)$$

References

1. Sinusoidal Analysis of Feedback-Control System Containing Nonlinear Elements, E. Calvin Johnson. *AIEE Transactions*, vol. 71, pt. II, July 1952, pp. 169–81.

2. Steady-State Axial Forces on Control Valve Pistons, Shih-Ying Lee, J. F. Blackburn. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 74, 1952, pp. 1005–11.

3. Berechnung von Ausfluss und Ueberfallzahlen, R. Von Mises. *Zeitschrift des Vereines Deutscher Ingenieure*, Düsseldorf, Germany, vol. 61, May-June 1917, pp. 447–52; 469–74; 493–98.

4. Generalized Charts of the Effects of Nonlinearities in Electrohydraulic Control Valves, J. Zaborszky, H. J. Harrington. *AIEE Transactions*, vol. 76, pt. I, May 1957, pp. 191–98.

5. A Method for the Selection of Valves and Power Pistons in Hydraulic Servos. F. C. Paddison, W. A. Good. *Report CM-717*, Applied Physics Laboratory, Johns Hopkins University, Baltimore, Md., Jan. 26, 1952.

Generalized Charts of the Effects of Nonlinearities in Electrohydraulic Control Valves

J. ZABORSZKY
MEMBER AIEE

H. J. HARRINGTON
ASSOCIATE MEMBER AIEE

THIS PAPER presents summary results of analog computer studies of electrohydraulic control valves. All principal nonlinearities associated with this type of control valve as well as nonlinearities common to actuators and inertia-type loads were instrumented on the analog computer.

The results are presented in a generalized, nondimensional form and cover a rather wide range of parameters. This makes the results applicable to valves of arbitrary dimensions, permits the drawing of certain generalized conclusions and serves as a basis for comparison with describing function studies.[1,2] The information presented herein should also prove useful in creating a better understanding of the dynamic performance of electrohydraulic servo systems and in the establishment of certain objectives to be achieved in the design of such systems.

The System Studied

The study refers to a 4-port single-stage electrohydraulic valve of the type shown in Fig. 1 which is driving an inertia load through a hydraulic actuator. Nonlinearities of the valve considered are the quadratic flow versus pressure drop relationship, the reaction or Bernoulli forces[3,4] acting on the valve spindle as a result of the flow, dashpot type of damping of the valve spindle (Fig. 1), coulomb friction on the valve spindle and torque motor and, in some cases, hysteresis in the torque motor.

The actuator piston is assumed to have coulomb friction and the oil in the actuator cylinder and oil ducts is assumed to be compressible. It is also assumed that the oil volumes connected to either side of the actuator piston are constant and equal; this assumption is justified as long as the oscillations are of small amplitude as considered here. The inertia-type load is assumed to be connected through an elastic coupling to the actuator piston and to have viscous and coulomb friction and a linear restoring force.

There is an electrical feedback through a lag network from the actuator piston, Fig. 2, as well as an inner loop resulting from counterelectromotive force induced in the torque motor by the movement of its armature and the valve spindle.

The equations which describe the system and which were instrumented on the analog computer are established in the Appendix.

Scope of the Study

The system under consideration has a tendency to oscillate at a frequency near the natural frequency of the load. This tendency can be established on the basis of a purely linearized representation of the system. However, the gain at which the oscillation starts is influenced to a very large extent by nonlinearities. These nonlinearities also determine the amplitude and influence the frequency of the oscillation and are responsible for certain peculiarities. For instance, in most cases when the gain rises above the minimum necessary to maintain the oscillation, the amplitude of the oscillation can be forced to jump abruptly from zero to a definite, relatively large value. Yet this oscillation cannot start spontaneously; it must be induced by a sufficiently vigorous disturbance.

Another possible mode of oscillation is attributable to the coulomb friction of the valve spindle. This also is a nonlinear oscillation but it occurs at a frequency which is different (smaller in the present study) from the frequency of the load. A third mode of oscillation occurs around the linear natural frequency of the valve spindle. This oscillation is likely to be at a high frequency.

Most of the study was devoted to measuring the minimum or drop-out gain, that is, the gain at which the sustained oscillation starts or stops, as well as to measuring the amplitude and frequency of the sustained oscillation. The mode of oscillation near the resonant frequency of the load was studied predominantly; this mode is of particular interest because a large number of nonlinearities simultaneously exert substantial effects on it. Other nonlinear modes of this kind of system tend to be established essentially by one nonlinearity only.

The effect of variations of various linear and nonlinear parameters are summarized in the curves which shall be discussed in the sections that follow. In general only one parameter was varied at a time while the other linear and nonlinear parameters were kept at their nominal values as given in a list following the nomenclature. In some cases, however, as many as three parameters were varied over a certain range.

Nomenclature

A = cross-sectional area of actuator piston
A_0 = cross-sectional area of damping orifice
A_s = cross-sectional area of valve spindle
$B = 2C_qK_a$ cos 69 degrees, reaction (Bernoulli) force coefficient[3,4]
C_q = discharge coefficient of valve metering orifices
E = error emf (electromotive force) appearing at servo amplifier input
F_{ac} = coulomb friction force acting between actuator piston and cylinder
F_c = coulomb friction force on valve spindle
$F_d = \dfrac{A_s^3}{2K_s^2A_0^2}$ orifice-damping coefficient of valve spindle
F_{lc} = coulomb friction force on load
f_a = viscous damping coefficient of actuator piston relative to cylinder
f_l = viscous damping coefficient of load
f_m = viscous damping coefficient of valve spindle and torque motor
$h = \dfrac{AP}{M_aX_{00}\omega_n^2}$, dimensionless form of A
I = differential current in torque motor
I_h = half-width of torque-motor-current hysteresis loop
$i = \dfrac{IK_m}{M_mX_{e0}\omega_n^2}$, dimensionless form of I
$K = \dfrac{1}{rq}\dfrac{y_{la}^2+y_b^2}{y_{la}^2+\left(1+\dfrac{nw^2}{h}\right)y_l^2}$, total open-loop gain for linearized[2] system
K_a = peripheral width of metering orifices in valve
K_f = steady-state gain of feedback network, i.e., $\dfrac{\Gamma_{a0}}{X_{00}}$
K_l = spring constant of restraining force on load
K_{la} = spring constant of linkage between actuator piston and load
K_m = force constant of torque motor referred to valve spindle

---

Paper 57-108, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 11, 1956; made available for printing November 26, 1956.

J. ZABORSZKY is with Washington University, St. Louis, Mo., and is consultant to McDonnell Aircraft Corporation, St. Louis, Mo.; H. J. HARRINGTON is with McDonnell Aircraft Corporation, St. Louis, Mo.

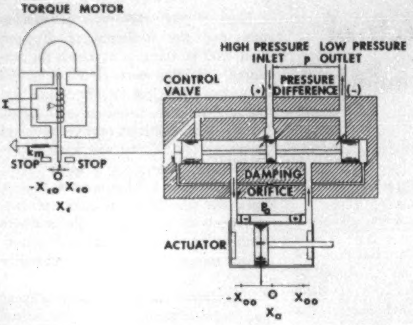


Fig. 1. Sketch of single-stage 4-port electrohydraulic control valve and torque motor

$K_v=\sqrt{1/\rho}$
K_e = coefficient of back emf induced by motion of torque-motor armature
$k=\frac{X_{00}}{X_{l0}}$, dimensionless proportionality factor
k_m = spring constant of torque motor referred to the valve spindle
$k_e=\frac{K_m K_e}{K_m L\omega_n^2}$, dimensionless form of K_e
L = inductance of torque motor winding
$l=\frac{K_m\Gamma_{a0}\mu}{LM_m X_{\epsilon 0}\omega_n^2}$, dimensionless form of μ
M_a = mass of actuator piston
M_l = mass of load
M_m = mass of valve spindle plus equivalent mass of torque-motor armature
$n=\frac{V_0 P}{4\beta A X_{00}}$, dimensionless oil compressibility parameter
P = accumulator-sump pressure differential; this constant also represents the maximum value of P_a
P_a = pressure differential: pressure in right chamber minus pressure in left chamber, Fig. 1, across actuator piston
$p=\frac{P_a}{P}$, dimensionless form of P_a
$q=\frac{\omega_n A X_{00}}{T\sqrt{P}X_{\epsilon_0}}$, dimensionless flow coefficient of valve-metering orifices
R = resistance of torque motor winding
$r=\frac{R+r_p}{L\omega_n}$, dimensionless form of $R+r_p$
r_p = plate resistance of output stage of amplifier
$T=C_q K_v K_a$
t = time
$u=\frac{\omega}{\omega_n}$, dimensionless frequency of the fundamental harmonic of a periodic oscillation
V_0 = total volume of oil on both sides of the actuator piston between the piston and the valve
$v=\frac{BP}{k_m+BP}$, dimensionless reaction-spring nonlinearity ratio
$w=\sqrt{\frac{K_{la}}{M_a\omega_n^2}}$, dimensionless form of the undamped natural frequency of the actuator piston
X_a = true actuator piston displacement for compressible oil; positive displacement from center to right; see Fig. 1
X_i = input emf to servo
X_l = displacement of load
X_{l0} = maximum displacement of load from center
X_0 = idealized actuator piston displacement for incompressible oil and no oil-ring backlash
X_{00} = maximum actuator-piston displacement from center
X_ϵ = valve spindle displacement; positive to right; see Fig. 1
$X_{\epsilon 0}$ = maximum valve spindle displacement from center
$x_a=\frac{X_a}{X_{00}}$, dimensionless form of X_a
$x_i=\frac{X_i}{\Gamma_{a0}}$, dimensionless form of X_i
$x_l=\frac{X_1}{X_{l0}}$, dimensionless form of X_1
$x_0=\frac{X_0}{X_{00}}$, dimensionless form of X_0
$x_\epsilon=\frac{X_\epsilon}{X_{\epsilon 0}}$, dimensionless form of X_ϵ
$y_l=\sqrt{\frac{K_1}{M_1\omega_n^2}}$ dimensionless form of the undamped natural frequency of the load referred to K_l when K_{la} is assumed zero
$y_{la}=\sqrt{\frac{K_{la}}{M_l\omega_n^2}}$, dimensionless form of the undamped natural frequency of the load referred to K_{la} when K_l is assumed zero
$\alpha\Delta$ = time constant of lag portion of feedback network
$\alpha\delta=\alpha\Delta\omega_n$, dimensionless time constant of lag portion of feedback network
β = bulk modulus of oil
Γ_a = output emf of feedback network
Γ_{a0} = maximum steady-state output emf of feedback network corresponding to $X_a=X_{00}$
$\gamma_a=\frac{\Gamma_a}{\Gamma_{a0}}$, dimensionless form of Γ_a
Δ = time constant of lead portion of feedback network
$\delta=\Delta\omega_n$, dimensionless time constant of lead portion of feedback network
$\epsilon=\frac{E}{\Gamma_{a0}}$, dimensionless form of E
$\zeta=\frac{f_m}{2\sqrt{M_m(k_m+BP)}}$, dimensionless form of f_m
$\zeta_a=\frac{f_a}{2\sqrt{M_a K_{la}}}$, dimensionless form of f_a
$\zeta_{ac}=\frac{F_{ac}}{2M_a X_{00}\omega_n^2}$, dimensionless form F_{ac}
$\zeta_c=\frac{F_c}{2M_m X_{\epsilon 0}\omega_n^2}$, dimensionless form of F_c
$\zeta_d=\frac{F_d X_{\epsilon 0}}{2M_m}$, dimensionless form of F_d
$\zeta_h=\frac{I_h K_m}{2M_m X_{\epsilon 0}\omega_n^2}$, dimensionless form of I_h
$\zeta_l=\frac{f_l}{2\sqrt{M_l K_{la}}}$, dimensionless form of f_l
$\zeta_{lc}=\frac{F_{lc}}{2M_l X_{l0}\omega_n^2}$, dimensionless form of F_{lc}
μ = gain of servo amplifier
ρ = oil density
$\tau=\omega_n t$, time measured in radians on basis of ω_n
ψ = peak amplitude of the fundamental harmonic of a periodic oscillation in p
ω = frequency, in radians per unit time, of the fundamental harmonic of a periodic oscillation
$\omega_N=\sqrt{\frac{BP}{M_m}}$, equivalent natural frequency of valve spindle when referred to the nonlinear reaction force
$\omega_n=\sqrt{\frac{k_m+BP}{M_m}}$, undamped natural frequency of the linearized valve[1,2]
$'$, $''$, primes denote derivatives with respect to τ

Any consistent system of units may be used.

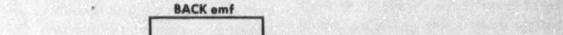


Fig. 2. Block diagram of closed-loop control system incorporating electrohydraulic control valve and actuator

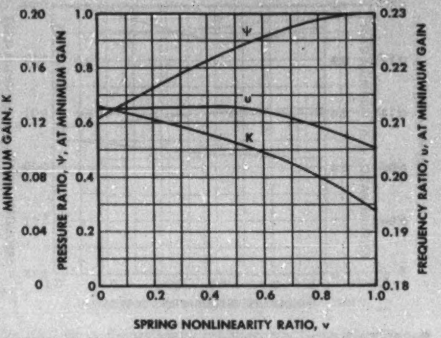


Fig. 3. Effect of spring nonlinearity ratio v on minimum gain K, required to sustain oscillations, on frequency u and pressure-amplitude ψ of these oscillations. All linear and nonlinear parameters except v have values as given in section listing parameters

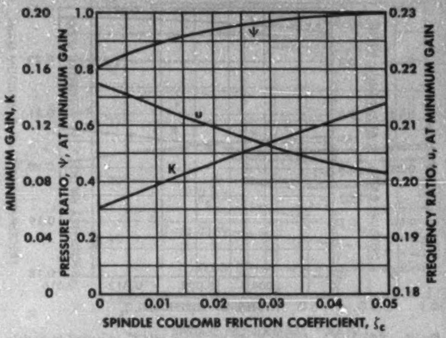


Fig. 4. Effect of coulomb friction of valve spindle ζ_c on minimum gain K which is required to sustain oscillations, on frequency u and pressure-amplitude ψ of these oscillations. All linear and nonlinear parameters except ζ_c have values as given in section listing parameters

Nominal Dimensionless Parameters of System Studied

$h=0.366$
$K=1.00$
$k=1.00$
$k_s=0.102$
$l=2.53$
$n=0.0106$
$q=68.5$
$r=2.14$
$v=0.705$
$w^2=10.12$
$y_l^2=0.00375$
$y_{la}^2=0.0573$
$\delta=0$
$\zeta=0.0441$
$\zeta_a=0.522$
$\zeta_{ac}=0$
$\zeta_c=0.0159$
$\zeta_d=6.30$
$\zeta_h=0$
$\zeta_l=0.0332$
$\zeta_{lc}=1.55\times10^{-6}$

Various Effects

Effect of reaction (Bernoulli) forces on valve spindle. The parameter v expresses what portion of the total spring constant acting on the valve spindle is derived from the reaction or Bernoulli forces caused by the flow through the valve with zero load on the actuator piston.[3,4] This reaction force is proportional to the valve opening; in this it acts as a linear spring. However, it is also proportional to the pressure drop across the valve. Since the pressure drop across the valve is a function of the instantaneous load motion, the reaction force becomes a nonlinear feedback which produces a very substantial destabilizing effect, as shown by Fig. 3. In Fig. 3, K represents the drop-out gain which is that value of gain just sufficient to sustain a limit-cycle oscillation. The quantities ψ and u represent the nondimensional peak amplitude and frequency respectively of the fundamental harmonic of the periodic oscillation in the pressure drop across the actuator p at the gain K. Fig. 3 indicates that the gain margin is reduced by more than half when the fluid reaction force increases from v = 0.0% to v = 100% of the spring constant.

Studies based on the steady operation of the valve indicate that a relatively high value of v such as v = 0.7 may be desirable for better flow-rate linearity in the steady state. This study now reveals that the increased linearity in the steady state is purchased at the price of reduced dynamic stability.

The effect of coulomb friction on the valve spindle is shown in Fig. 4. As may be expected, increasing coulomb friction improves stability as far as oscillation at the natural frequency of the load is concerned; doubling the coulomb friction increases the stability margin by roughly 15% to 25%.

On the other hand, coulomb friction is also responsible for nonlinear oscillations in another mode at much lower frequencies and much higher gains (in the system studied) than the mode near the natural frequency of the load. The nature of these latter oscillations is highly dependent upon the lag-filter characteristics in the main feedback path. When the filter time constant was zero, no such oscillations could be detected for reasonable loop gains.

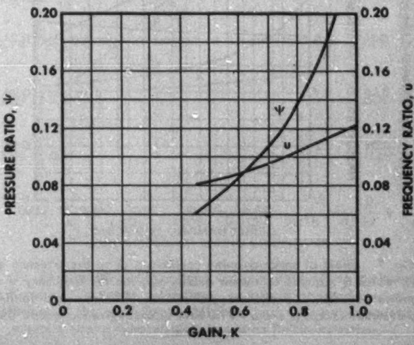


Fig. 5. Pressure-amplitude ψ and frequency u of a coulomb-friction-type sustained oscillation as functions of open loop gain K

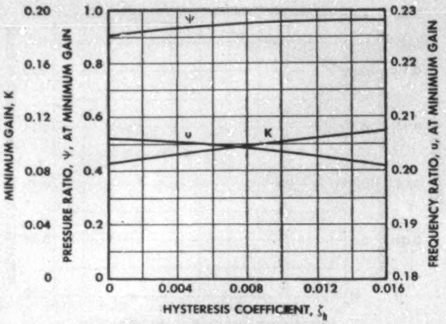


Fig. 6. Effect of hysteresis ζ_h on minimum gain K, required to sustain oscillations; on the frequency u and pressure-amplitude ψ of these oscillations. All linear and nonlinear parameters except ζ_h have values as given in section listing parameters

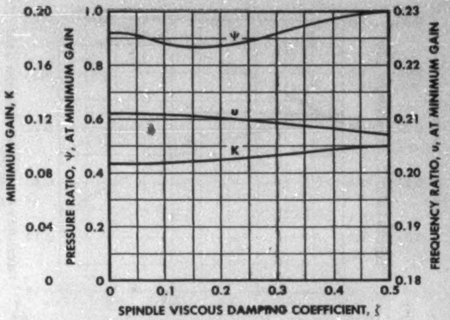


Fig. 8. Effect of viscous damping ζ of the spindle on the minimum gain K required to sustain oscillations; on the frequency u and pressure-amplitude ψ of these oscillations. All linear and nonlinear parameters except ζ have values as given in section listing parameters

Some results for the latter type of oscillation are shown in Fig. 5 for a single value ($\zeta_e = 0.0159$) of the Coulomb friction of the spindle. This figure shows the variation with gain of the frequency and pressure ratio of the sustained oscillation. For this mode, no clear-cut dropout gain was found.

Effect of hysteresis in torque motor or backlash between torque motor and valve spindle. The analog computer representation was that of hysteresis with a rhomboid-shaped characteristic. However, backlash tends to have identical effects unless it is associated with appreciable bouncing or inertia phenomena.

The effect of hysteresis is also similar to that of coulomb friction. Like coulomb friction, hysteresis or backlash also have a moderately stabilizing influence at the mode of oscillation coincident approximately to the resonant frequency of the load, Fig. 6; all are capable of introducing additional modes of oscillation at other frequencies.

The similarity between hysteresis and coulomb friction can readily be appreciated by comparing Figs. 4 and 6. The two curves become essentially coincidental by letting $\zeta_e = (\zeta_h + 0.0159)$ considering that $\zeta_e = 0.0159$ is present in Fig. 6.

Effect of dashpot-type damping on valve spindle. As shown in Fig. 7, the effect of dashpot-type damping (Fig. 1) is not as pronounced as that of coulomb friction at the mode of oscillation associated with the natural frequency of the load. Its effect becomes very pronounced, however, at the higher frequency mode associated with the natural frequency of the valve spindle itself.

Effect of viscous friction on valve spindle. Viscous friction has a slight stabilizing effect and exercises a small influence on the amplitude and frequency of oscillations around the resonant frequency of the load, as shown in Fig. 8.

Effect of compressibility of oil. Oil compressibility for small oscillations essentially represents a linear spring located between the valve and actuator. Increasing the oil compressibility coefficient *n*, therefore, effectively softens a

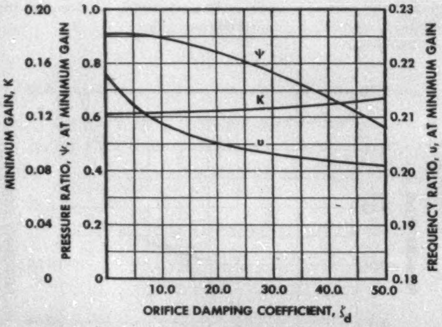


Fig. 7. Effect of orifice-damping coefficient, ζ_d on the minimum gain K, which is required to sustain oscillations; on the frequency u and pressure-amplitude ψ of these oscillations. All linear and nonlinear parameters except ζ_d and q = 97.0 have values as given in section listing parameters

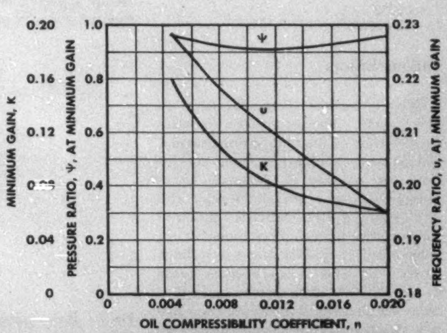


Fig. 9. Effect of oil compressibility n on minimum gain K required to sustain oscillations; on frequency u and pressure-amplitude ψ, of these oscillations. All linear and nonlinear parameters except n have values as given in section listing parameters

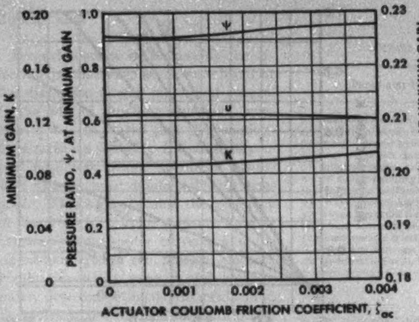


Fig. 10. Effect of coulomb friction at actuator ζ_{ac} on minimum gain K required to sustain oscillations; on frequency u and pressure-amplitude ψ of the oscillations. All linear and nonlinear parameters except ζ_{ac} have values as given in section listing parameters

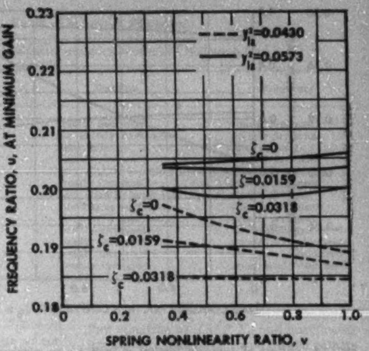


Fig. 12. Joint effect of spring nonlinearity ratio v of coulomb friction in valve ζ_c and of stiffness of coupling between actuator and load y_{la}^2, on frequency u of oscillation sustained by minimum gain K. All linear and nonlinear parameters except v, ζ_c, y_{la}^2, and $y_l^2=0$ have values as given in section listing parameters

spring within the closed loop. The effect, as might be expected, is a substantial reduction of frequency and an even more substantial reduction of stability margin (Fig. 9). It is noteworthy that as the compressibility becomes larger the rate of change in the gain margin is gradually decreasing.

Effect of coulomb friction in actuator. For the oscillation at the resonant frequency of the load, coulomb friction in the actuator has a stabilizing effect which is rather insignificant according to Fig. 10, at least within the investigated range.

Effect of varying load natural frequency with two other parameters. Summarized in Figs. 11 and 12 is the effect of a joint variation of three parameters:

1. The spring constant between actuator and load inertias, y_{la}^2.
2. That portion of the total spring constant on the valve derived from the reaction forces v.
3. Coulomb friction on the valve spindle.

In contrast to the spring constant which is represented by the compressibility of the oil and is located inside the loop, increasing the compressibility (decreasing the spring constant y_{la}^2) of the spring between the actuator and load inertias improves stability since this latter spring is outside of the loop.

Increasing values of v, as has already been pointed out, Fig. 3, have a destabilizing effect. This effect is now observed to be slightly more pronounced if the coulomb friction on the valve spindle ζ_c is increased or if the spring between the actuator and load inertia y_{la} is softer.

The effect of coulomb friction ζ_c on

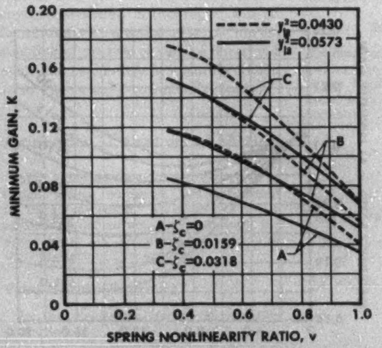


Fig. 11. Joint effect of spring nonlinearity ratio v of coulomb friction in valve ζ_c and of stiffness of coupling between actuator and load y_{la}^2 on minimum gain K required to sustain oscillations. All linear and nonlinear parameters except v, ζ_c, y_{la}^2, and $y_l^2=0$ have values as given in section listing parameters

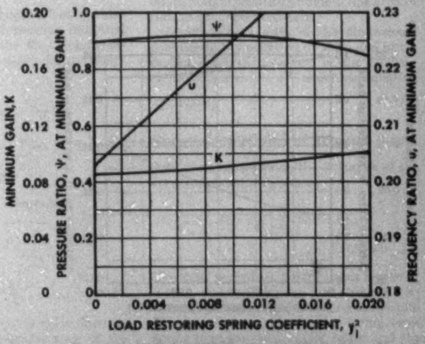


Fig. 13. Effect of stiffness of restoring spring on load y_l^2 on minimum gain K required to sustain oscillations; on frequency u and pressure-amplitude ψ of the oscillations. All linear and nonlinear parameters except y_l^2 have values as given in section listing parameters

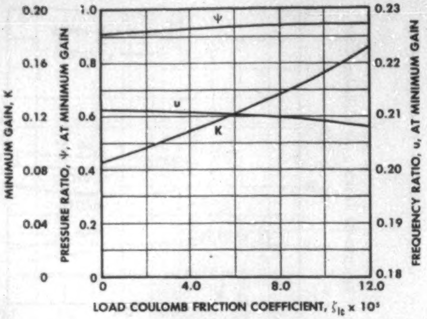


Fig. 14. Effect of coulomb friction at load ζ_{lc} on minimum gain K required to sustain oscillations; on frequency u and pressure-amplitude ψ of these oscillations. All linear and nonlinear parameters except ζ_{lc} have values as given in section listing parameters

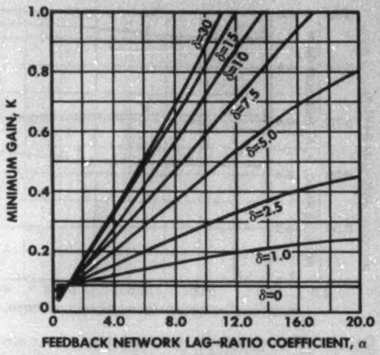


Fig. 16. Joint effect of feedback network lag-ratio α and of feedback network time constant δ on minimum gain K required to sustain oscillations. All linear and nonlinear parameters except α and δ have values as given in section listing parameter

the valve spindle was found, Fig. 4, to be stabilizing. Now, from Fig. 11 it is seen to be slightly more stabilizing if v is large or if the spring between the actuator and load inertia is harder.

It is interesting that the relatively slight effect of both v and ζ_e on the frequency of the oscillation, Fig. 12, tends to reverse for the two values of y_{la} studied. Of course, the effect of y_{la} itself is much more pronounced on the frequency than that of v or ζ_e since these oscillations develop around the natural frequency of the load.

Effect of restoring spring constant of load. The effect of the restoring spring on the load appears principally in the frequency of the oscillation and to a lesser extent in the gain margin where a harder spring produces more stability. In these respects the spring is similar to that resulting from the compressibility of the oil; see Fig. 13.

Effect of coulomb friction on load. As shown by Fig. 14, coulomb friction acting on the load has a rather substantial stabilizing effect on the system when the oscillation is around the resonant frequency of the load. However, oscillation at other modes may be produced by this coulomb friction.

Effect of viscous friction on load. As might be expected, viscous friction on the load effectively stabilizes the system while it has only a minor effect on the frequency and amplitude of the oscillation; see Fig. 15.

Effect of feedback network. A lag network in the feedback path very substantially influences the system performance, as shown in Figs. 16 and 17. Generally, an increase of both the lag-ratio constant α and the time constant, δ improves stability; the influence of α being strongest when δ is relatively large, and vice versa; (Fig. 15). This network can have a substantial effect even on the frequency of the oscillation; see Fig. 17.

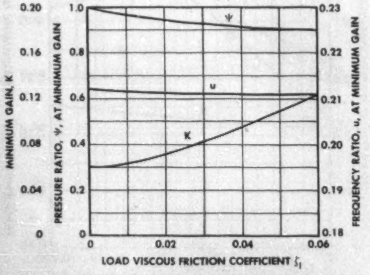


Fig. 15. Effect of viscous friction at load ζ_l on the minimum gain K required to sustain oscillations; on the frequency u and pressure-amplitude ψ of these oscillations. All linear and nonlinear parameters except ζ_l have values as given in section listing parameters

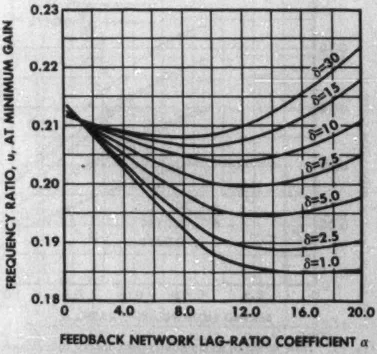


Fig. 17. Joint effect of feedback network lag-ratio α and of feedback network time constant δ on frequency u of oscillation maintained by minimum gain K. All linear and nonlinear parameters except α and δ have values as given in section listing parameters

Conclusions

Several conclusions concerning the effects of parameter changes were listed in the discussion of the individual curves. In the way of general conclusions, however, it can be stated that the reaction-force ratio v is probably the most important nonlinear design parameter associated with electrohydraulic valve performance. This is true because the constancy of flow rate with valve pressure drop at constant torque-motor current, and the closed-loop system gain margin, are both highly dependent upon v and unfortunately pose conflicting requirements on it. In addition, the reaction-force ratio represents the only nonlinear valve-parameter studied which may be easily varied and controlled by proper design procedures.[3,5]

The importance of keeping the total volume of fluid V_0 between the actuator and the valve at a minimum in order to minimize the oil compressibility parameter n is also evident from this study. This linear parameter should also be minimized if a high resonant frequency is desired, as is usually the case.

The effect of a numerical increase in each of the parameters studied is qualitatively evaluated in Table I.

The fact that only one parameter at a time was varied in obtaining most of the results presented appears not to be a very serious shortcoming, since in the one case (Figs. 11 and 12) involving the variation of three parameters simultaneously, little interdependence was found between the effects of the individual parameters.

It is believed that the data and discussions presented in this report will be found useful in creating a better understanding of the role of various linear and nonlinear parameters in electrohydraulic servo valves and representative loads.

Table I. Result of Increasing Parameter

Results Apply to Load-Resonant-Frequency Mode of Oscillation Only

Parameter Increased	Stability	Frequency
v = reaction-force ratio	greatly decreased	almost constant
ζ_c = spindle coulomb friction	moderately increased	slightly decreased
ζ_h = torque motor hysteresis	same as for ζ_c	same as for ζ_c
ζ_d = spindle orifice damping	slightly increased	slightly decreased
ζ = spindle viscous friction	very slightly increased	slightly decreased
n = oil compressibility	greatly decreased	greatly decreased
ζ_{ac} = actuator coulomb friction	very slightly increased	almost constant
y_{la}^2 = actuator-to-load spring rate	moderately decreased	slightly increased
y_l^2 = load restoring force spring rate	moderately increased	greatly increased
ζ_{lc} = load coulomb friction	greatly increased	moderately decreased
ζ_l = load viscous friction	greatly increased	very slightly decreased
α = feedback network lag-to-lead ratio	greatly increased	moderately decreased
δ = feedback network lead time constant	greatly increased	moderately decreased

Appendix. Equations of Motion

The following equations were used to design analog computer instrumentation for the servo system. To promote a better appreciation of an actual system, both the dimensional and nondimensional forms of the equations will be presented.

Valve Spindle Dynamics

In the dimensional form, the equation of motion is

$$(I\pm I_h)K_m=M_m\ddot{X}_e+f_m\dot{X}_e+F_c\frac{\dot{X}_e}{|\dot{X}_e|}+F_d\dot{X}_e|\dot{X}_e|+k_mX_e+B\left(P+P_a\frac{X_e}{|X_e|}\right)X_e \qquad (1)$$

Here the left-hand term represents the force, distorted by hysteresis I_h, delivered to the valve spindle by the torque motor. On the right-hand side the first term is the inertia force; the second, third, and fourth terms represent viscous, coulomb friction, and orifice (dashpot) damping, respectively; the fifth term is the force derived from the mechanical spring situated in the torque motor; and the last term represents the total reaction (Bernoulli) force acting on the spindle.[3,4]

Combining the linear portion of the reaction force with the mechanical spring force and dividing both sides of the dimensional equation by M_mX_{e0} leads to the intermediate equation of

$$\frac{(I\pm I_h)K_m}{M_mX_{e0}}=\frac{\ddot{X}_e}{X_{e0}}+\frac{f_m}{M_m}\frac{\dot{X}_e}{X_{e0}}+\frac{F_c}{M_mX_{e0}}\frac{\dot{X}_e}{|\dot{X}_e|}+\frac{F_dX_{e0}}{M_m}\frac{\dot{X}_e}{X_{e0}}\frac{|\dot{X}_e|}{X_{e0}}+\frac{k_m+BP}{M_m}\frac{X_e}{X_{e0}}+\frac{BP}{M_m}\frac{P_a}{P}\frac{|X_e|}{X_{e0}} \qquad (2)$$

By introducing symbolism defined in the Nomenclature, this may be written as

$$(i\pm 2\zeta_h)\omega_n^2=\ddot{x}_e+2\zeta\omega_n\dot{x}_e+2\zeta_c\omega_n^2\frac{\dot{x}_e}{|\dot{x}_e|}+2\zeta_d\dot{x}_e|\dot{x}_e|+\omega_n^2x_e+\omega_N^2p|x_e| \qquad (3)$$

Dividing both sides of equation 3 by ω_n^2 and making the time transformation indicated in the Nomenclature, the final nondimensional form of the valve-spindle dynamics becomes

$$(i\pm 2\zeta_h)=x_e''+2\zeta x_e'+2\zeta_c\frac{x_e'}{|x_e'|}+2\zeta_dx_e'|x_e'|+x_e+vp|x_e| \qquad (4)$$

Flow at Valve Orifices

The dimensional flow equation is

$$A\dot{X}_0=C_qK_eK_aX_e\sqrt{P+P_a\frac{X_e}{|X_e|}}=TX_e\sqrt{P+P_a\frac{X_e}{|X_e|}} \qquad (5)$$

which simply expresses the fact that flow rate is proportional to the square root of the pressure drop across the valve orifices and also to the area K_aX_e of the orifices. The oil is assumed incompressible while being metered by the valve.

By dividing both sides of the dimensional flow equation by $T\sqrt{P}\,X_{e0}$ and introducing appropriate changes of variables and parameters, the equation may be nondimensionalized. Thus

$$qx_0'=x_e\sqrt{1+p\frac{x_e}{|x_e|}} \qquad (6)$$

Compressibility of Oil

The oil is assumed to be compressible when considered in the region between the valve and the actuator piston. This is necessary because of the relatively large volume of oil in the actuator cylinder and in the tubing. In dimensional form the equation expressing oil compressibility is

$$A(X_a-X_0)=\frac{V_0}{4\beta}P_a \qquad (7)$$

The term on the left represents the change in oil volume due to the difference in position of the actuator piston when the oil is considered incompressible, X_0, and compressible, X_a. The term on the right is the same change of volume expressed by the oil compressibility. Note that the assumption is made that the oil volume is equal and constant on either side of the actuator piston. This is a justified assumption for oscillations of small amplitude acting around zero as are considered here.

Dividing both sides of the dimensional compressibility equation by AX_{a0} and introducing symbolism defined in the Nomenclature leads to the dimensionless equation

$$x_a-x_0=np \qquad (8)$$

Equation of Motion of Actuator Piston

In dimensional form, the dynamics of the actuator piston are given by

$$M_a\ddot{X}_a+f_a\dot{X}_a+F_{ac}\frac{\dot{X}_a}{|\dot{X}_a|}+K_{la}(X_a-X_l)+AP_a=0 \qquad (9)$$

The first, second, and third terms represent inertia, viscous damping, and coulomb friction forces respectively acting on the piston. The fourth term represents a spring force acting between the actuator piston and the load. The last term is of course the external force applied to the piston by the oil.

When both sides of the dimensional piston equation are divided by $M_aX_{a0}\omega_n^2$ the equation reduces to the dimensionless form

$$x_a''+2\zeta_a w x_a'+2\zeta_{ac}\frac{x_a'}{|x_a'|}+ w^2\left(x_a-\frac{1}{k}x_l\right)+hp=0 \quad (10)$$

Equation of Motion of Load

The load assumed for this study is described by the dimensional equation

$$M_l\ddot{X}_l+f_l\dot{X}_l+F_{lc}\frac{\dot{X}_l}{|\dot{X}_l|}+K_{la}(X_l-X_a)+ K_lX_l=0 \quad (11)$$

Here, the first four terms are similar to those described for the actuator piston while the fifth represents a linear restraining force acting on the load.

By dividing both sides of the load equation by $M_lX_{l0}\omega_n^2$ and introducing new notation, the following dimensionless equation may be found:

$$x_l''+2\zeta_l y_{la}x_l'+2\zeta_{lc}\frac{x_1'}{|x_1'|}+ y_{la}(x_l-kx_a)+y_l^2x_l=0 \quad (12)$$

Equation of Feedback Network

The electrical network in the feedback path is assumed to have the dimensional transfer function of $\Gamma_a/X_a=K_f(\Delta s+1)/(\alpha\Delta s+1)$ for which the differential equation is

$$\alpha\Delta\dot{\Gamma}_a+\Gamma_a=K_f(\Delta\dot{X}_a+X_a) \quad (13)$$

In dimensionless notation, this becomes the following

$$\alpha\delta\gamma_a'+\gamma_a=(\delta x_a'+x_a) \quad (14)$$

Equation of Error Signal

Dimensionally, the error signal is given by

$$E=X_l-\Gamma_a \quad (15)$$

Dividing both sides by Γ_{a0}, this equation becomes

$$\epsilon=x_l-\gamma_a \quad (16)$$

Differential Equation of Amplifier and Torque Motor

The differential equation of the amplifier and torque motor is, in dimensional form the equation

$$\mu E=L\dot{I}+(R+r_p)I+K_e\dot{X}_e \quad (17)$$

where the left-hand term represents the amplified error signal with a perfectly flat amplifier characteristic and the first two terms on the right symbolize the inductance-resistance electrical circuit comprised of the amplifier output stage and the torque-motor winding. The last term represents the counterelectromotive force produced by the movement of the torque-motor armature.

Multiplying both sides of the dimensional equation by $K_m/LM_mX_{e0}\omega_n^2$ leads to the dimensionless equation

$$l\epsilon=i'+ri+k_ex_e' \quad (18)$$

References

1. A Describing Function for Electrohydraulic Control Valves, J. Zaborszky, H. J. Harrington. *Report no. 4580*, McDonnell Aircraft Corporation, St. Louis, Mo., Feb. 20, 1956.

2. A Describing Function for Multiple Nonlinearities Present in Electrohydraulic Control Valves, J. Zaborszky, H. J. Harrington. *AIEE Transactions*, vol. 76, pt. I, May 1957, pp. 183–90.

3. Steady-State Axial Forces on Control Valve Pistons, Shih-Ying Lee, J. F. Blackburn. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 74, 1952, pp. 1005–11.

4. Berechnung von Ausfluss und Ueberfallzahlen, R. Von Mises. *Zeitschrift des Vereines Deutscher Ingenieure*, Dusseldorf, Germany, vol. 61, May-June 1917, pp. 447–52; 469–74; 493–98.

5. A Method for the Selection of Valves and Power Pistons in Hydraulic Servos, F. C. Paddison, W. A. Good. *Report CM-717*, Applied Physics Laboratory, Johns Hopkins University, Baltimore, Md., Jan. 26, 1952.

An Inductronic Double Bridge

J. H. MILLER
FELLOW AIEE

THE Kelvin or double bridge is considered as the standard type of bridge for the measurement of very low resistance values and, more specifically, where 4-terminal resistances are used. Typical examples are in the comparison of shunts which are basically 4-terminal resistors and in the measurement of conductivity where the current terminals are at the ends of the wire or bar under consideration, and the millivolt terminals are between those current terminals.[1–3]

The Kelvin bridge involves somewhat more elaborate circuitry than the simple Wheatstone bridge although the form using multiple-deck dial switches somewhat simplifies the manipulation. Nevertheless, this general arrangement appears necessary where the bridge current must flow through low-value resistors in series as two adjacent bridge arms, and which have an anomalous resistance between them. Frequently the voltage drop between the two resistors is larger than the drop in the resistors themselves and essentially the Kelvin bridge provides an appropriate return point for the galvanometer.

In the adjustment of shunts, the Kelvin bridge has been used rather consistently. However, the manipulation required along with appropriate maintenance of the equipment is not as simple as is desirable in the manufacturing plant.

Development of the Method

When a shunt or other low-valued resistor is to be adjusted in terms of the resistance of a standard, it is obviously possible to pass a current through the two resistors in series and compare the voltage drop across them. Theoretically, if the ratio of these drops could be obtained directly, the resistance of the test shunt in terms of the standard could at once be determined. But, since the two voltages to be compared are not only small but are also separated by a larger anomalous voltage, it would appear that they can only be compared in the Kelvin network or in a 2-coil ratiometer of rather limited sensitivity.

To simplify the measurement procedure in the production of such low-valued resistors or shunts, consideration was then given to the possibility of amplifying the millivolt values, using d-c amplifiers where there is no internal connection from the input to the output terminals and which would allow for the amplified millivolt values to be compared in an accurate manner, perhaps in an interconnected network. Amplifiers embodying a mechanical link in the amplifying chain would function in this way as would those where the signal is converted to an alternating current, amplified with transformer coupling and phase-rectified back to direct current or potential. Where such amplifier systems operate

---

Paper 57-111, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted July 2, 1956; made available for printing December 12, 1956.

J. H. Miller is with the Weston Electrical Instrument Corporation, Newark, N. J.

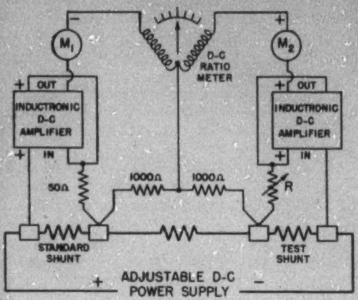


Fig. 1 (left). Inductronic double bridge with indicating ratiometer

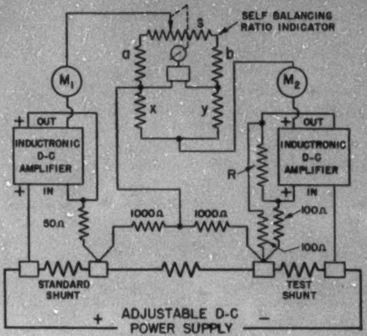

SELF BALANCING RATIO INDICATOR
S
a
b
x
y
M1
M2
OUT
INDUCTRONIC D-C AMPLIFIER
IN
R
50Ω
1000Ω
1000Ω
100Ω
100Ω
STANDARD SHUNT
TEST SHUNT
ADJUSTABLE D-C POWER SUPPLY
+
–

Fig. 2 (right). Inductronic double bridge with self-balancing ratio indicator of limited range

with full feedback through a mutual resistor, very high accuracy of gain is possible.

The inductronic system described by Gilbert[4] is of this type, where a d-c input is converted into alternating current, amplified, phase-rectified, and, in the usual form, the output is fed back through a mutual resistor to completely balance the input; the amplifier, in essence, only amplifies the unbalance or error component. As applied to the present problem, one feature of the inductronic system is that, with a millivolt input and a proportional milliampere output, that output will remain truly proportional to the input without regard to the resistance loading in the current output circuit, if the *IR* drop in that output circuit is less than 5 volts. Since less than 5 volts would be used across the low-valued resistors in series, it is thus possible to tie the two outputs together in any of several types of ratio indicators.

As with many other d-c amplifiers, the current output of the inductronic amplifier is somewhat limited and normally considered as 1.0 ma (milliampere) maximum. This is a rather small current value for a pivot-type ratiometer, so that the preliminary study envisioned the use of a servo-type ratiometer where the ratio of the two currents, balanced against a ratio obtained from a motor-driven slide wire, resulted in a difference signal, which, being amplified, balances the motor. This is essentially the mechanism of a typical ratio recorder.

The study having progressed thus far, it was also considered advisable to determine if the shunt under test could be matched to a standard shunt having some other rating; when balanced, unity being indicated on the ratiometer, the test shunt would be correct. This apparently could be arranged by adjusting the feedback resistance in one of the amplifiers.

Circuit Details

The first circuit which evolved is shown in Fig. 1. The 50-ohm resistor at the left is the amplifier feedback or coupling resistor and is of good accuracy. *R* is the feedback resistor in the test shunt amplifier. The two 1,000-ohm resistors bridging the potential terminals of the two shunts simply serve as return circuits for the amplifier output currents; they are nominally alike since, with unity ratio of these currents, the anomalous voltage across them to be absorbed in the amplifier output circuits is evenly divided. Lower values can be used, and the circuit will function correctly even if one of these units is reduced to zero. It is important, where multiple leads to shunt potential connections are shown, that these be completely independent up to the potential terminals. In practice, the several leads are soldered in adjacent holes in a relatively heavy copper lug, clamped to the potential terminal on the shunt.

When 50 mv (millivolt) appears across the standard shunt and the associated amplifier, it causes the output current to rise until the *IR* drop of this current through the feedback resistor balances the 50-mv value. Thus, on applying the 50-mv input, the output current in M_1 rises to 1.0 ma and holds this value. The transconductance of the amplifier with a 50-ohm coupling resistor could be expressed as 20,000 micromhos, or the reciprocal of 50 ohms.

Similarly, in the amplifier network connected to the test shunt, for 1.0-ma output to balance 50 mv, *R* must be 50 ohms and this would be the circuit condition if the test shunt were exactly equal to the standard; the ratiometer would then indicate unity. If the test shunt is 10% low in resistance, giving 45 mv under the same conditions, 0.9 ma is developed in the output and the ratiometer will indicate 0.9, or that the test shunt is 10% low in resistance. Presumably the shunt would then be adjusted to unity on the ratiometer.

Alternatively, if the test shunt has twice the current rating of the standard

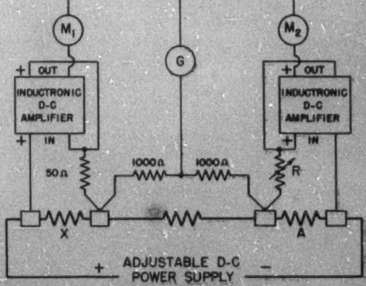


Fig. 3. Inductronic double bridge with galvanometer balance indicator; at balance $X = 50\ A/R$ ohms

shunt (for 50-mv drop) or half of the resistance, then, with current to produce 50 mv in the standard shunt, there is only 25 mv in the test shunt. To get a balance or a unity-ration indication that the test shunt is correctly adjusted, R becomes 25 ohms so that the 1.0-ma output produces 25 mv through R. In other words, at unity ratio or balance of the output current of the two amplifiers,

$$\frac{\text{test shunt resistance, ohms}}{\text{standard shunt resistance, ohms}}=\frac{R}{50} \qquad (1)$$

Then, for a unity indication of the ratiometer showing the test shunt to be correctly adjusted to its rated value,

$$R=50\ (\text{test shunt, ohms})/(\text{standard shunt, ohms}) \qquad (2)$$

Alternatively, in terms of shunt rating in amperes for the same millivolt values,

$$\frac{\text{standard shunt rating}}{\text{test shunt rating}}=\frac{R}{50} \qquad (3)$$

or

$$R=50\ (\text{standard shunt rating})/(\text{test shunt rating}) \qquad (4)$$

Since, as stated in the foregoing, a direct reading d-c ratiometer operating at 1.0 ma has a limited sensitivity, a self-balancing ratio indicator was added to the system as shown in Fig. 2. A somewhat limited range was selected because shunts are fabricated from the excellent manganin available today, rolled to a nominal thickness, and held close to a specific resistance per square. Shunts as fabricated are from 1% to 5% low in resistance before adjustment. A span of 10% seemed appropriate for the ratio indicator, and the final scale selected was from 92% to 102% of nominal. With a 100-ohm slide wire for the 10% span, resistors a and b are 920 and 980 ohms respectively. Exact slide-wire values are difficult to achieve, however, and are not necessary here; any value from 50 to 200 ohms might be used. For a 10% span covering 92% to 102%, with a slide wire of s ohms, $a=9.2s$ and $b=9.8s$. The ratio arms, x and y are usually 500 ohms each and, while their exact values are not important, they must be equal to each other to better than 0.02%. To arrange for the checking of a shunt having some ratio of other than unity to the standard shunt, the network at the right in Fig. 2, the coupling resistor, must be variable. Because this requires contacts, the network was changed as shown. Here the potential in millivolts, balancing the test-shunt millivolt drop, is

$$i_{ma}\left(\frac{100\times100}{R+200}\right) \qquad (5)$$

For 1.0-ma output current, it is $10{,}000/(R+200)$. To balance 50 mv, R then is zero and this will balance the 50-ohm coupling resistor associated with the standard shunt. This particular arrangement assumes that the test shunt will always have a rating higher than the standard shunt, a rather desirable procedure in practice. At an output unity current ratio, from equation 3

$$\frac{\text{standard shunt rating}}{\text{test shunt rating}}=\frac{10{,}000/(R+200)}{50}=\frac{200}{R+200} \qquad (6)$$

or

$$R=200\left(\frac{\text{test shunt rating}}{\text{standard shunt rating}}-1\right)\text{ ohms} \qquad (7)$$

Thus, for the same current rating $R=0$; while for a test shunt rated at twice the current of the standard shunt for the same millivolt drop, $R=200$ ohms. With standard shunts or resistors in decade steps, R may have a maximum value of 2,000 ohms; the smallest rating ratio will be 1.2, where $R=40$ ohms. Values of R in this network are thus large enough so that contact resistance up to 0.01 ohm can be disregarded.

For a practical installation, the monitoring instruments M_1 and M_2 indicate on their 1.0-ma scales the output current of the amplifiers, and for best accuracy the main current is adjusted until approximately 1.0 ma exists in the output. This is important since these amplifiers fail to be proportional because of overloading at about 2 ma, and in the same way, with a given resolution, best accuracy is obtained with an adequate input level.

Conclusions

Equipment for checking shunts as described has been built and functions in a satisfactory manner; accuracy is excellent. If all power is disconnected and the pointer of the ratio indicator is manually moved to a different position, when current is again applied the previous value is indicated to within 0.05% on the scale. Shunts adjusted using this equipment have been found to check on a conventional Kelvin bridge to better that 0.1%. As a practical matter, the simplicity of operation of the measuring gear has resulted in more accurate shunt adjustment even with the same allowed tolerances.

It was pointed out previously that the results are independent of the loading of the amplifier output if the drop is not over 5 volts; to prove this, as much as 6,000 ohms were arbitrarily added to one of the output circuits before any change in the indication became evident.

Although the full 1.0 ma in the output of the amplifiers is desirable, with the available resolution in the amplifiers 0.1% accuracy can still be maintained at as low as 0.1 nominal rated current (0.1 ma in the amplifier output), which allows for adjusting and checking shunts rated at 50,000 amperes and using a 5,000-ampere power supply.

For use as a resistance-measuring bridge, Fig. 3 is suggested with a decade box substituted for R of Fig. 1, and a null galvanometer G used as a balance indicator. As R is adjusted to indicate null, and with a 50-ohm coupling resistor as shown in the left-hand amplifier, then $X=50\,A/R$.

This general arrangement appears to be a rather powerful method for measuring low resistances in a straightforward and direct manner with a good degree of accuracy. The resolution of the amplifiers is less than 5 microvolts; if sufficient current is put through the resistor under test and through the standard to produce about 50 mv in each, the basic resolution is one part in 10,000. Taking into account all other factors and assuming the important resistors are adjusted to 0.02%, it seems readily possible to obtain an over-all accuracy of 0.1%

This application would appear to stress again the importance of the use of the full feedback amplifier as a link in many of our classical measuring methods.

References

1. ELECTRICAL MEASUREMENTS (book), F. K. Harris. John Wiley & Sons, Inc., New York, N. Y., 1952, pp. 282–89.

2. ELECTRICAL MEASUREMENTS (book), F. A. Laws. McGraw-Hill Book Company, Inc., New York, N. Y., 1938, pp. 186–93.

3. BASIC ELECTRICAL MEASUREMENTS, M. B. Stout. Prentice-Hall Publishing Company, New York, N. Y., 1950, pp. 80–84.

4. THE INDUCTION GALVANOMETER, A SENSITIVE INSTRUMENT CONVERTER, R. W. Gilbert. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1121–26.

General Aspects of Hands-Free Telephony

J. W. EMLING
MEMBER AIEE

Synopsis: The use of a simple hands-free telephone set with distant talking and loudspeaker reception introduces transmission degradation not encountered with hand-set usage. Proper usage and good acoustic surroundings reduce these impairments to a point where satisfactory transmission is obtained on a large percentage of calls in today's plant which is designed to have transmission margin on a good proportion of all calls. There are several potentialities for improving the hands-free telephone set but it is not yet clear whether or not the gain will justify the added complexity or the added restrictions placed on the user. It is to be expected that very excellent performance could be obtained from a unit such as a booth which combines optimum instrument design with favorable environment.

IT IS perhaps natural that a demand for hands-free telephony should arise in this era of hands-free dishwashing, hands-free gear shifting, and the multitude of other available labor-saving devices. Like many labor-saving devices, however, hands-free telephony presents problems not existing with more conventional approaches. These problems, particularly as they affect the transmission engineer, are discussed in this paper.

From a transmission standpoint, the fundamental difference between the hands-free telephone and the conventional handset is in the distance between the instruments and the user; see Fig. 1. As the instruments are moved away from the head, a loss is introduced in the transmitting and receiving paths amounting, in each direction for common hands-free arrangements, to some 20 db (decibels). In addition, as the distance is increased the characteristics of the room in which the set is used begin to play an important part. With the handset, the instruments are so close to the head that performance is largely independent of the surrounding conditions, but with distant instruments room conditions can be controlling.

To recover the distance loss it is necessary to introduce amplification in the transmitting and receiving paths, but this is limited by the loudspeaker-microphone feedback paths. Some of the speech entering the microphone is reproduced by the loudspeaker as side tone, and this speech in turn travels through the air path to the microphone. At all frequencies, the total amplification in this loop must be kept below the total loss, or "singing" will occur. The losses are affected not only by the side-tone balance at the set but also very appreciably by the room characteristics. Speech from the loudspeaker travels to the microphone both by a direct path, as shown in Fig. 1, and by way of reflection from walls, the body, furniture, etc., as in Fig. 2. The aggregate of these reflected sounds, at a single frequency at least, can be much greater than the directly transmitted sound.

This is illustrated by Fig. 3 which shows as a function of room characteristics, the amount of loop gain which can be introduced before singing occurs. The point of reference is a perfect room, i.e., one completely devoid of reflection. The room constant R* is used to define the reflection characteristics of other rooms.[1] For any given distance from a sound source, the ratio of directly received energy to the total reverberant energy is a function of this constant. For free space (perfect room), there is no reverberation and R is infinity. A large office with rug, drapes, and sound-absorbing ceiling will have a constant of 1,000 or more, and an 8- by 10-foot room with hard surfaces will have one as low as 200.

Referring to Fig. 3, it is seen that, for the specific conditions illustrated, the introduction of a user together with a table to support the hands-free telephone reduced the usable gain in an otherwise perfect room by about 8 db, and another 10- to 12-db reduction occurred as room reflection was increased. Thus a small room may require as much as a 20-db reduction in gain as compared with free space.

It is apparent that there is a definite limitation in amplification depending on the surrounding conditions and, because the gain is limited, it is important to use it to best advantage. If all gain is put in the receiving branch the user receives best performance, but a large burden is thrown on the distant party who receives none of the benefits of hands-free telephony. If all gain is introduced in the transmitting path, high levels can be transmitted to the distant party but the hands-free user will be unduly handicapped. The solution used in the Bell System has been to introduce enough transmitting gain to provide transmitted levels from a properly used hands-free telephone about equal to those from a handset on long loop. Receiving gain is under control of the user and the amount usable depends on room characteristics, being largest for rooms with good acoustic characteristics and smallest for those which are highly reverberant. Thus the user's reception is dependent upon the acoustic conditions under which he chooses to use his telephone.

In addition to limiting the total transmitting-receiving gain, the room conditions and the side-tone balance affect the quality of the speech transmitted. Initially, consider only the room effects and assume, for the moment, that the side-tone path is nonexistent. When the transmitter is at some distance from the lips, the sound which actuates the transmitter reaches it by several paths. One, and the most important, is the direct line from the lips to the transmitter. However, the sound also spreads out though the room and is reflected back from the walls, furniture, and other solid objects as shown in Fig. 4. These reflected sounds are always present in a room, but at a level much lower than the sound leaving the lips of the speaker. When a handset is used, the sound received directly from the lips is at a very high level and the reflected sounds are so relatively low that the listener at the other end of the circuit is not aware of their presence. When the transmitter is removed to a considerable distance from the lips, the direct sound reaching it is so reduced that the reflected sounds approach it in magnitude. The listener is very conscious of their presence particularly since the reflected sounds, having traveled around the room, arrive at the transmitter after the direct sounds

---

Paper 57-137, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted December 19, 1956; made available for printing January 3, 1957.

J. W. EMLING is with the Bell Telephone Laboratories, Inc., New York, N. Y.

The author wishes to acknowledge the assistance of David Nast in pointing out the negative-impedance aspects of the problem, Harry Neues in obtaining the data shown on Figs. 6 through 9, and the late Norman Stryker for his studies of singing and reverberation.

* $R = \alpha S_R/1-\alpha$, where α is the average absorption coefficient of the room surfaces, and S_R is the surface area.

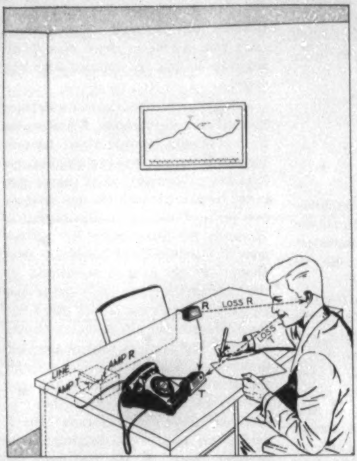


Fig. 1. Direct transmission path with hands-free telephone

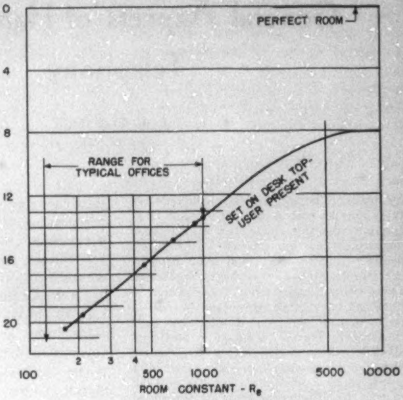


Fig. 3. Singing versus acoustical environment

and thus produce a hollow effect, commonly called reverberation, which is very much like talking into a barrel.

The reflected sounds are present to the same extent when talking through the air to a listener in a reverberant room. The reverberation is much less annoying, however, since the listener in this case uses 2-ear, or binaural, reception which makes it possible for him to concentrate on the sound which reaches his ear first. When a telephone circuit is used, the listener effectively employs only a single ear, placed at the microphone position, and he can no longer exclude the effects of the delayed reflected sounds.

With a distant talking telephone, the annoyance of reverberation depends upon the distance between the lips and the transmitter, and also upon the characteristics of the room, such as its size, shape, and the amount of sound absorbed by the structure. This is shown by Fig. 5, which indicates the relation between the room constant and the distance between lips and transmitter at which reverberation becomes noticeable, i.e., the point at which the reflected energy is 10 db less than the incident energy. Note that the talking distance should be kept under about 18 inches if reverberation effects in typical offices are to be kept negligible. As an illustration of room reverberation, Fig. 6 shows oscillo-

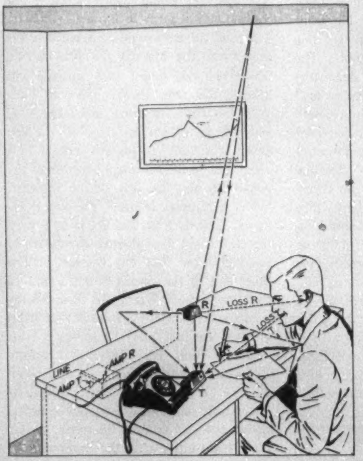


Fig. 2 (left). Indirect transmission between loudspeaker and microphone

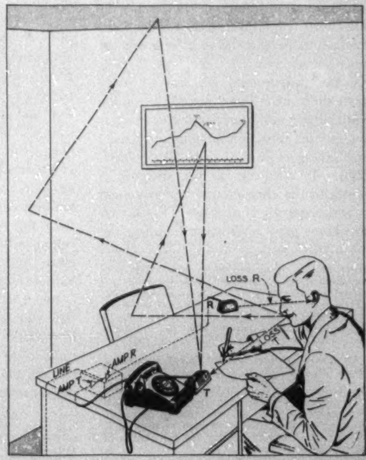


Fig. 4 (right). Reflection giving rise to reverberation

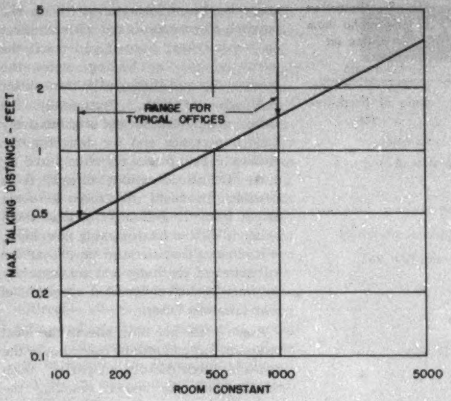


Fig. 5. Maximum talking distance without reverberant effects

grams of the line voltage at the set terminals, produced by acoustic pulses introduced into a handset and hands-free set.* Even though the acoustic path is only 18 inches for the latter, the effects of reverberation are noticeably greater than with a handset. Addition of amplification in the transmitting path will not improve the situation since both direct and reverberant energy will be increased accordingly.

It is apparent, therefore, that the reverberation effects with a hands-free set can be kept small only by keeping the path loss from mouth to transmitter small, and by providing surroundings which minimize the reflected energy. The surroundings are beyond control of the set designer unless they are an integral part of an enclosure, such as a booth, but the path loss can be minimized by providing arrangements which encourage close talking.

While low path loss and low reflectivity in the room are essential to obtain speech with little reverberation distortion, they do not ensure it since the effects of side-tone speech are much the same as reflected speech. These sounds are appreciably delayed, particularly in a highly reverberant room, and also can be comparable in magnitude to the direct speech. Side tone, therefore, can provide a reverberant effect even with otherwise ideal acoustic surroundings if the gain in the path is sufficient to cause near singing conditions. The oscillograms of Fig. 7 illustrate the important part which side-tone can play.

The reverberant effect, while produced at the hands-free set, is noticeable only at the distant end. Toll transmission engineers will recognize it as a form of multiple listener echo. Another effect, noticeable at the distant end, is talker echo. It is seen in Fig. 8(A) that incoming speech at the hands-free set travels from the loudspeaker by direct and reflected paths to the microphone whence it is returned to the distant talker with a delay depending on the round-trip time of transmission over the wire lines and the delay in the acoustic path at the hands-free telephone set. The latter amounts to about 3 milliseconds as a minimum, for 3-foot separation between the loudspeaker and microphone, and may be several times as long for reflected sounds.

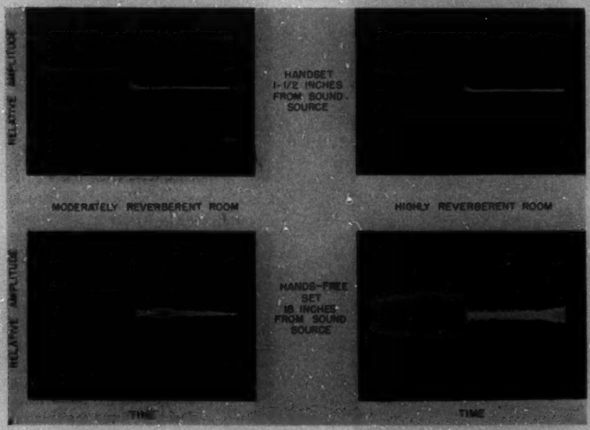


Fig. 6. Oscillograms of voltage at set terminals from 10-microsecond pulses of 1,000-cycle tone, side-tone path open

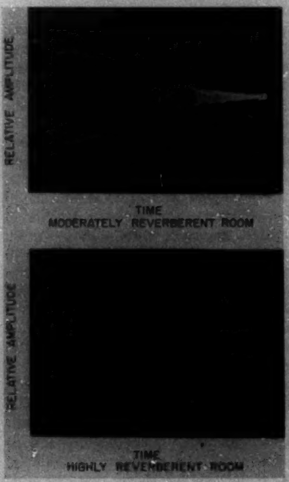


Fig. 7. Oscillograms as in Fig. 6 but with side-tone path gain adjusted for near singing, direct and side-tone wave approximately in phase

---

* A pulse of single frequency is obviously an inadequate simulation of a complex speech wave, but it does provide a simple illustration of some important acoustic effects.

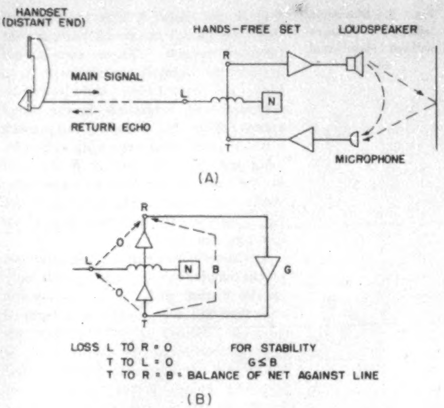


Fig. 8. Illustration of talker echo from hands-free set

A—Echo path
B—Equivalent schematic of hands-free set

The reflected wave may be considerably greater than the received wave. This is illustrated by the oscillograms of Fig. 9. The reason can be seen quite readily if the actual circuit is translated into the equivalent shown in Fig. 8(B). Here, small gains are introduced in the transmit and receive paths to just cancel the division and dissipation losses of the hybrid coil, and the gain G represents the sum of the losses and gains from R to T by way of the loudspeaker, air path, and microphone. The losses from L to R and from T to L will now be zero, and the reflected wave at point L (or T) will be greater than the received wave at L (or R) by the gain G. This gain must be less than the balance B of the network against the line but can approach it at a single frequency and consequently can be sizeable with good side-tone balance. If, for example, the quite good average balance of 20 db is achieved and the average margin against singing is 8 db, then G will be 12 db and the reflected wave at L will be 12 db higher than the received wave. Toll transmission engineers will recognize such a set as a line termination with a negative return loss amounting, in this example, to 12 db. They will also recognize the seriousness of such a termination on a toll circuit with appreciable delay.

This is an admittedly "rough-and-ready" analysis. There are a number of other possible approaches. For example, K. S. Johnson has shown that any anti-side-tone circuit with receiver coupled to transmitter is equivalent to a *21*-type repeater. For such a repeater the input impedance becomes negative with a large amplifying ratio and approaches $-Z$ net. Hence the hands-free set, when operated near singing and with a good side-tone balance, provides a negative impedance termination to the line approaching as a limit the line impedance with 180-degree shift in phase.

While echo and singing are the effects most noticeable on calls involving hands-free sets, another factor, room noise, may also be of importance. With the handset, room noise is not without effect on transmission, since it can reach the telephone ear by leaking under the receiver cap and through the transmitter and side-tone path. Fortunately, the noise reaching the free ear is of relatively little importance and by designing the receiver cap to fit the ear closely and by using the anti-side-tone circuit, it is possible to shield the user to some extent from the full effect of local room noise. With a loudspeaking telephone, it is obvious that the user must bear the full brunt of the noise and consequently may need a higher received speech level than a handset user.

Room noise not only affects the local reception but can also be carried over the line and affect the distant party. With the handset, the speech reaching the transmitter, due to its proximity to the lips, is much higher than the room noise reaching the transmitter, and the latter is seldom important since the local noise at the receiving end is usually higher. With a distant-talking transmitter, however, the speech reaching the transmitter is much lower, and room noise and speech levels are more nearly the same. Thus noise from the room in which the distant-talking set is used may be noticeable to the far-end listener. Here again the situation is not improved by increasing

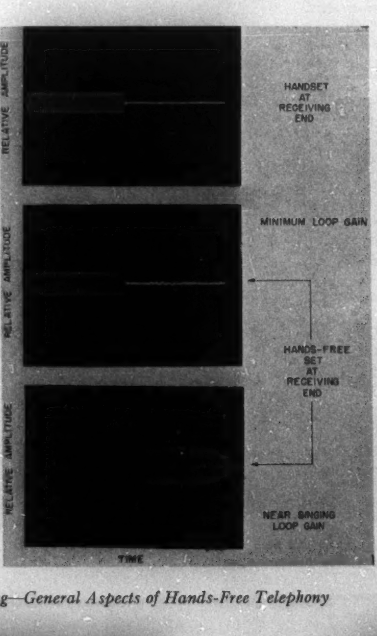


Fig. 9. Oscillograms of voltage at receiving-set terminals from 10-microsecond pulses of 1,000-cycle tone, introduced at distant (transmitting) end of circuit. Received and reflected waves approximately in phase

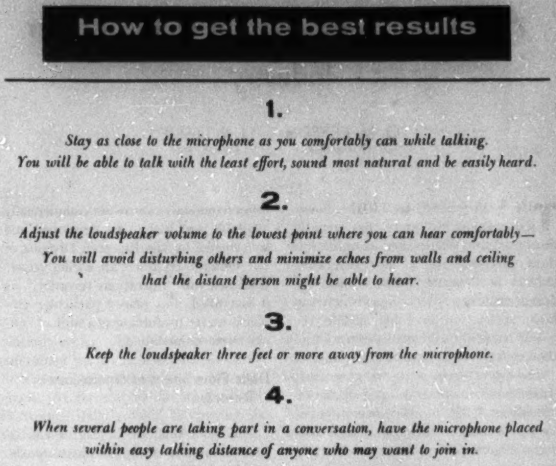
transmitting amplification since this will increase the transmitted room noise as well as the speech. The proper corrective measures are the reduction of room noise, which is mainly out of Telephone Company control, and the reduction of the acoustic loss between lips and transmitter.

In summary, it is seen that a hands-free telephone set affects not only the transmission received by the user but also that received by the distant party as well. The user may be affected by singing or alternatively by undesirably low received volume, particularly if he has noisy surroundings. At the distant end, the noticeable effects are listener and talker echo and perhaps transmitted room noise. The importance of these effects can be reduced by proper use of the set and by its use in a favorable acoustic environment.

Proper use requires adequate separation between loudspeaker and microphone with the minimum practical distance between the instruments and the user. The former increases the singing margin and the latter minimizes the gain required in the feedback loop. A favorable environment is one with low room noise which avoids transmitted noise effects and minimizes the required receiving volume. The room should also be reasonably free from reflection to minimize reverberation and increase singing margin, or usable gain.

Although environmental factors are beyond the control of the designer, unless the set is incorporated in a booth, an understanding of the proper usage conditions on the part of the user and the salesman will greatly increase the effective use of the device. Fig. 10, taken from a recent Bell System pamphlet, is an interesting example of how education in proper use can be coupled with merchandizing.

As indicated in another paper,[2] observance of the usage rules together with the favorable acoustic surroundings, which are becoming common in business locations, leads to successful use in a high percentage of cases. No small reason for this is that today's telephone plant is designed so that transmission on most calls exceeds the minimum requirements of good communication by a considerable margin. As a result, a large percentage of all telephone calls can tolerate the moderate degradations involved.



How to get the best results

1.

Stay as close to the microphone as you comfortably can while talking. You will be able to talk with the least effort, sound most natural and be easily heard.

2.

Adjust the loudspeaker volume to the lowest point where you can hear comfortably— You will avoid disturbing others and minimize echoes from walls and ceiling that the distant person might be able to hear.

3.

Keep the loudspeaker three feet or more away from the microphone.

4.

When several people are taking part in a conversation, have the microphone placed within easy talking distance of anyone who may want to join in.

Fig. 10. Use of education with merchandizing

While the main objective of this paper was to outline the problems which are inherent in hands-free telephony, it may not be out of place to discuss briefly design measures which might aid in their solution. It has already been pointed out that the problems can be greatly ameliorated by reducing the path loss between the instruments, particularly the microphone, and the user. This can theoretically be accomplished by using directive instruments. However, achieving significant directivity in a highly reverberant room, where it is most needed, is not an easy matter. Even if this is achieved, it is not certain that the spacial restriction placed on the user will be acceptable. Reduction of the distance between mouth and microphone will also reduce the path loss and can theoretically be accomplished by providing microphones (or loudspeakers) on elevated arms, clipped to the clothing, or attached to a headband as in an operator's telephone set. However, all of these curtail the user's freedom to some extent, and it is doubtful that the customer would feel that the benefits offset the restrictions placed on him.

Considerable improvement is also possible, theoretically, by the use of voice-operated devices. Since talker and listener echoes returned through the loudspeaker-microphone path are an important source of difficulty, it would be logical to solve this problem with echo-suppressor techniques. Specifically, this would require a voice-operated device which would open (or insert loss in) the receiving branch while talking and perform the reverse operation during reception. This gain change could also be introduced in an inverse manner in the transmitting branch. It remains to be seen whether or not the complexities and other effects such as clipping will be justified by the benefits obtained.

Finally, much can be accomplished if the instruments and their environment can be designed and supplied together in the form of a booth on which distances, noise, and reflection can be closely controlled.

References

1. A LOUDNESS RATING FOR LOUDSPEAKERS AND POWER REQUIREMENTS FOR ROOMS, H. F. Hopkins, N. R. Stryker. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 36, Mar. 1948, pp. 315–35.

2. THE BELL SYSTEM SPEAKERPHONE, W. F. Clemency, F. F. Romanow, A. F. Rose. *AIEE Transactions*, vol. 76, pt. I, May 1957, pp. 148–53.

Advanced Data System in the Power Industry

G. M. KEYSER
ASSOCIATE MEMBER AIEE

J. R. LESLIE
MEMBER AIEE

THE HYDRO-ELECTRIC Power Commission of Ontario has recently developed and constructed an automatic data processor for a new generating station at Niagara Falls. Since the design employs new techniques which may have application in fields outside the power industry, a detailed description of the key features is offered in this paper.

The operator in a power generating station must maintain a log of all station operations such as the operation of circuit breakers, protection relays, and manual controllers. These need attention when the operator is busy with his first duty of maintaining power output. After the station regains its normal steady state following system trouble, the operator must recall to the best of his knowledge what has happened, when it happened, and what steps he took as control measures.

As stations grew more complex the operator frequently fell behind in the logging. The quantity of data could exceed the level of efficient recording and vital information could be missed. Engineering analysis of the logged record was a slow and costly procedure. It involved supposition and theorizing before a conclusion could be reached and acted upon.

The employment of a second operator was not the answer, because of considerations of economy and because one operator would not necessarily know what had, or had not, been done by the other operator. It was decided that an automatic data processor was necessary, a processor which would record all data that had previously been recorded manually, plus the time or sequence of every operation.

The recording of every operation involved the problem of those that occurred simultaneously. Data on these occasions would have to be retained temporarily in storage units. A scan system would pick up the data from these units and transmit it to a printer. The printer would provide the final record in a form similar to that encountered in manual processing. This record could enter the log book directly.

A suitable printer was found, but the other requirements were not commercially available and had to be developed and constructed by the Research Division of the Ontario Hydro. An earlier paper[1] discussed this "operations recorder," as it was called, and placed particular emphasis on its usefulness as a unit of control room equipment.

Data Flow Sheet of Operations Recorder

The block diagram of Fig. 1 can be read as a flow sheet to illustrate the stages of data handling between the initial operation and the final record.

The electrical indication of an operation is the opening or closing of a contact on any of 540 points monitored by the operations recorder. Thus, at the left on the block diagram are the "input contacts." These lead to the "relay storage unit," whose function it is to retain the data on the operation. The 540 relay storage units are all separate, and can receive information simultaneously and retain it indefinitely.

A "scan relay" associated with each one of the relay storage units is a part of a scan system. The scan system allows only one scan relay to close at a time. This is essential because the printer is limited to producing one recrod at a time. The scan relays transmit the data from the relay storage units to the coding.

The "name coding" is a group of contacts which are physically mounted on the 540 scan relays. When a scan relay is closed, the name coding contacts will cause the printer to print the name of that point. When it is open, there is no printing. As stated, when one scan relay is closed the other 539 are open.

"Time," as a separate unit, is continuously available and will be included with any printing. The "printer"[2] produces one line at a time, of up to 40 characters, on 8$^1/_2$-inch paper. Each line of printing describes one operation.

The maximum data-processing rate is ten operations per second. At this speed the full capacity of the storage units can be cleared in 54 seconds. Operations slower than ten per second are cleared immediately from the storage units and hence recorded in sequence. Operations faster than ten per second are recorded as being simultaneous. The record shows when the operations are simultaneous. The remaining sections offer a detailed description of the relay storage units and the scan system since they were the parts which presented particular design problems.

Relay Storage Unit to Hold Data

The storage unit must record both an opening and a closing of the input contact. It must distinguish between these two operations. In addition, the input must be disconnected after an operation, as a multiple operation would confuse the store. With these requirements in mind a search of the literature was made, but this failed to locate circuits having the desired response. The relay circuits to be described were developed in the research division of Ontario Hydro and are believed to be unique.

Telephone relays were used in the storage unit because they could run on the available 48- and 250-volt station batteries, and their cost per contact was lower than alternative methods. Other choices investigated were magnetic cores and cold cathode triggered glow tubes. Both appeared as possible alternatives but would have required more development to eliminate potentially troublesome performance.

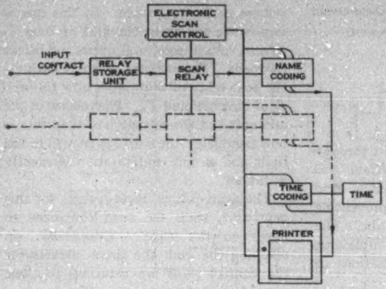
In addition to locking up one relay for either an opening or closing operation, the storage unit must indicate which type of operation has occurred. This is done with a second relay which will lock open for an opening operation, and lock closed for a closing operation. This provides two elements of data: the fact of the operation, and the nature of the operation. The relay circuit to perform these functions is shown in Fig. 2. The storage unit can be demonstrated by tracing the results of an operation leading to a definite configuration of locked-up relays.

Assume that the input contact, shown

Paper 57-221, recommended by the AIEE Recording and Controlling Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted July 10, 1956; made available for printing January 3, 1957.

G. M. KEYSER and J. R. LESLIE are with The Hydro-Electric Power Commission of Ontario, Toronto, Ont., Canada.

Credit is due to N. O. Ny for technical skill shown in the construction of the prototype and final models of the circuits described.



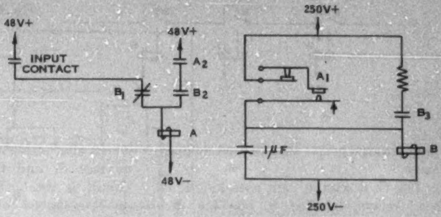


Fig. 1 (left). Block diagram of operations recorder

Fig. 2. Relay storage unit circuit

at the left of Fig. 2, performs an operation by closing. Relay A closes, due to current through contact B_1 and the relay coil. As relay A is closing, there is a momentary closure of springs a and b on contact A_1. These springs are open circuit when relay A is open or closed but closed momentarily during either the opening or the closing of the relay. In telephone practice, contact A_1 is known as a make-before-break assembly.

The momentary closure is the indication of an operation at the input contact. It causes relay B to lock up, and to store this information. Since the closure of contact A_1 is too brief to operate a relay directly, the 1-microfarad condenser is included. The condenser becomes fully charged in the available time, and discharges slowly through the relay. Relay B locks up through its own B_3 contact. The scan system is notified of an operation by B contacts additional to those shown in Fig. 2.

The nature of the operation is stored by locking relay A either open or closed, depending on the action of the input contact. Contacts B_1 and B_2 transfer the coil of relay A from the input contact to its own contact. A_2, which locks relay A open or closed. The case assumed was a closing, and relay A is locked closed by its closed A_2 contact.

The relays are now in their storage state. The closed relay B shows that there has been an operation, and the closed relay A shows that it was a closing operation. This information is transferred by the scan system to the printer for recording. After recording, the storage unit is returned to normal. This is accomplished by releasing the relays with a short circuit across the coil of relay B. This closes contact B_1, and the coil of relay A is again controlled by the input contact which may still be closed or may have reopened during the scan and printing cycle. The short circuit is maintained for 30 milliseconds to cover the release and operate times of both relays. At the end of the short circuit the storage unit is ready to respond to further operations.

Now, assume that the input contact performs an operation by opening. This causes relay A to open. As mentioned previously, when relay A opens there is a momentary closure of spring a and b on contact A_1, which leads to relay B being locked up. The contact B_2 transfers the coil circuit of relay A to contact A_2 which is open. This locks relay A open.

The configuration of the relays is again in its storage state. The closed relay B shows that there has been an operation, and the open relay A shows that it was an opening operation.

The foregoing has discussed the circuit operation of the relay storage. The physical details will now be given. Relays A and B, together with a third relay, are mounted on a plug-in chassis. The third relay is part of the scan circuit, and will be discussed separately. Fig. 3 shows a typical unit. The three relays are mounted with contacts accessible for cleaning and adjusting. The condenser and current-limiting resistors are mounted behind, and the complete unit plugs into a rack on the 32-pin connector at the back. The plated sheet metal chassis has been extended at the front as a dust shield and as a handle.

Locating Operated Relay Storage Units

Following each operation, a scan system is used to locate the operated storage unit for connection to the printer. The printer records only one operation at a time. When there are several simultaneous operations, the scan system must select one at a time for connection to the printer. A further requirement is equal opportunity for printing at each point.

The most direct scan system, and the one first tried in the development of the operations recorder, used stepping switches. The switch would step past each point, stopping and printing those holding operations. While this was a practical solution, it possessed the fundamental problem of waste motion in step-

Fig. 3. Relay storage unit

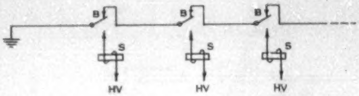


Fig. 4. Walking chain circuit

ping through points that carried no operation.

The final scan system uses individual scan relays, labelled *S*, one for each storage unit. Operated storage units are located directly with no waste motion. The circuit is a development of the walking chain circuit. Fig. 4 shows a simple walking chain circuit. When a relay *B* locks closed following an operation, it transfers its *B* contacts from the position shown in Fig. 4 to the coil circuit of relay *S*. This completes the circuit from high voltage *HV* to ground and closes the scan relay. The most important feature of this walking chain circuit is that it is immaterial whether the first or the 540th unit operated.

The final scan chain circuit is a further development of the one described, with features that eliminate certain problems. The first problem occurred when a unit operated while another unit, whose position was further along the chain, was printing. The *B* contact of the earlier unit interrupted the current to the printing scan relay, and closed the new scan relay. This interrupted printing was not acceptable, so a new chain of contacts was added to the original to hold each scan relay closed until printing was complete. The second problem occurs after the printing of one unit when operations exist at other units. In this case, the scan circuit must not favor those units earlier in the chain to the exclusion of the later ones. The problem was solved with two more chains which direct the subsequent locking up of the scan relays to the right. Only when there are no more operated units to the right will the earlier scan relays be given a second chance to print.

A full schematic of the final scan chain circuit is shown in Fig. 5. On comparison with Fig. 4, note that the *B* contacts, *S* relays, and tube T_1 form the circuit for the walking chain. The form of contact designation is changed to simplify the drawing.

The operation of the scan circuit will now be indicated by following simultaneous operations on unit 539 and unit 540. Fig. 5 shows all contacts in their inoperated state. On operation at unit 539, the *B* relay will close, providing a circuit through contact B_2, tube T_1, and relay *S*. As the *S* relay operates it sets up an alternative holding circuit through contact S_1 and tube T_2. When this circuit is established, the control grid on T_1 is moved to –40 volts, preventing further current through the tube. The circuit is now free of any troublesome effects due to subsequent operations of other units.

Printing and release of the stored information now occurs, with the *S* relay remaining closed. After the print-release cycle on unit 539 is completed, the scan cycle restarts to lock up the next operated unit to the right of unit 539. Bringing the control grid to T_3 to zero volts allows current flow through T_3, the closed S_4 contact of unit 539, the B_2 contact assumed closed because of the operation of unit 540, and the *S* coil of unit 540. Scan relay 540 will now operate but, unlike the first scan cycle that operated scan relay 539, an alternative holding circuit to T_4 by way of the S_3 contact of unit 539 is set up. The grid of T_3 is now returned to –40 volts and the circuits is independent of operations in other units, but there are now two scan relays closed.

To release the scan relay of unit 539 without releasing the scan relay of unit 540, it is necessary to change the grid of tube T_2 momentarily to –40 volts. This will stop the current to the scan relay of unit 539 and release that relay. Since this will eventually open the S_3 contact carrying the scan relay current for unit 540, it is important to establish another alternative circuit. This is done by requiring that contacts S_1, S_2, and S_3 release in that sequence. In this case it is possible to return the grid of tube T_2 to zero volts as soon as S_1 opens and before S_2 opens. The alternate current path for the scan relay of unit 540 is now through S_2 of unit 539 and T_2. The release of the one unit and the retention of the other is now complete; the scan has moved to the right and circuit configuration is exactly as before.

The print-release cycle repeats for this new unit, then the scan continues to move to the right. Eventually, on reaching the end, the drive circuits for the control grids are returned to their normal configuration as in Fig. 5. The scan cycle will continue as before if other operations have occurred during the printing of units 539 and 540, or they may wait in that state for the next operation. Regardless of the combination or timing of operations, the scan circuit has proved successful in sorting them out and feeding the information to the printer.

Electronic Circuits Program Scan Cycle

The chain of *S* relays is controlled by appropriate voltages on the grids of three tubes. The foregoing section has shown how the relays respond to current passed by the tubes. The following section discusses the electronic circuits that control the grid voltages on these tubes. The signals that control the electronic circuits are generated by the relay currents in the scan circuit, thus providing a closed-loop system that will program the scan cycle. A block diagram of the electronic scan control circuit is shown in Fig. 6.

The scan circuit of Fig. 5 will be recognized as a part of the over-all circuit. Each block is some form of

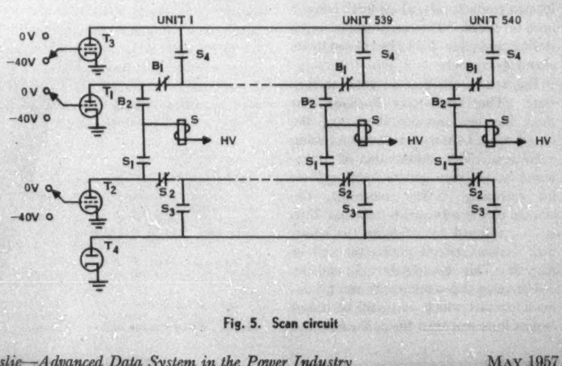


Fig. 5. Scan circuit

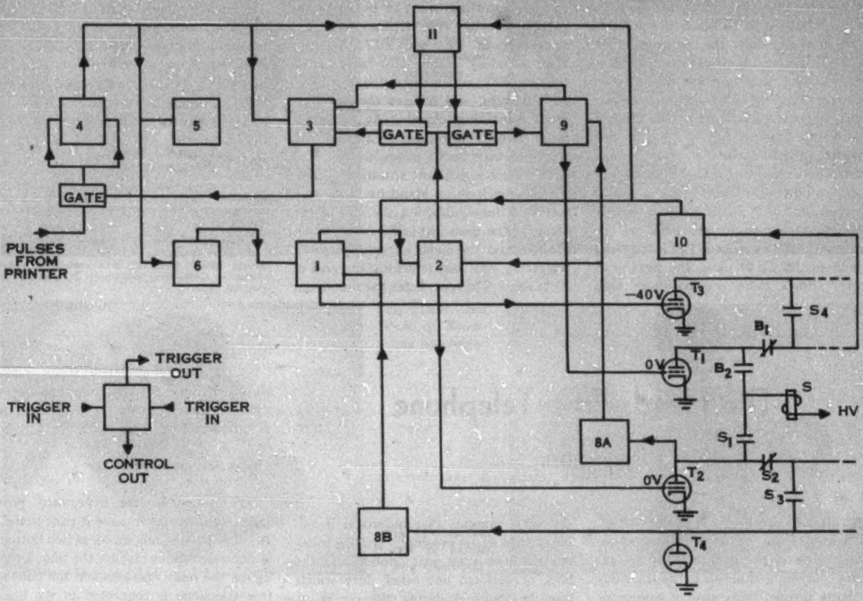


Fig. 6. Electronic scan control block diagram

flip-flop or multivibrator circuit having two characteristic states. The more common bistable flip-flop will stay in either state until triggered by external pulses. A valuable feature of the flip-flop is the large pulses produced as it changes state. The pulses are available in either polarity, and can be used to trigger further flip-flops.

The first operation in the scan cycle is to close one *S* relay and isolate all others so that they cannot interfere during the printing of the first. Refer now in Fig. 6, to the grid of tube T_1, where a connection can be noted leading to flip-flop 9 (*FF9*). This connection is made to a point on the flip-flop that is either at zero or −40 volts, depending on the state of that flip-flop. In the stand-by state the voltages are as shown on the grids of T_1, T_2, and T_3. An operation will lead to current flow through T_2 as explained earlier. At that time it was mentioned that the grid of T_1 was changed to −40 volts. The means of achieving this will now be explained. The current through T_2 is able to change the state of *FF8A* with an output pulse going to *FF9*. This pulse will change the state of *FF9* and change the grid of T_1 to −40 volts as required. The scan circuit now stands with one *S* relay closed, and will not be affected by subsequent operations of other relay storage units.

The scan cycle now continues with the printing of the information held in the one relay storage unit. The printing cycle uses four flip-flops and starts when *FF9* cuts off tube T_1. When *FF9* changes, it triggers *FF3*, which opens a gate. This gate receives a pulse from the printer each time the printer is prepared to record a new line of printing. The first pulse after the gate is open triggers *FF4*, which opens other gates for the coding pulses thus passing the signals for the printing of that operation. The next pulse from the printer returns *FF4* to normal and prevents further printing. An output pulse from *FF4* triggers four other flip-flops: *FF5* to advance the paper in the printer, *FF6* to release the storage unit relays, *FF11* setting it in a new state, and *FF3* returning it to normal. *FF6* returns to normal without a further trigger after 30 milliseconds, and sends an output pulse to *FF1*.

The following operation of the scan cycle is to lock up an *S* relay in the first operated unit to the right of the one just printed. The relay is closed with current through T_3 and held on with current through T_4. Similar to *FF9*, the grid connection of T_3 is controlled by *FF1*. When *FF1* is triggered, the grid of T_3 is raised from its normal −40 volts to zero volts, as mentioned in the earlier section on the scan circuit. This leads to the closing of the next scan relay with current through T_4.

The first scan relay must now be released while the other is held closed. This requires one more flip-flop plus a signalling circuit to indicate the release of the relay. The current through T_4, which follows the closing of the new relay, triggers *FF8B* which gives the pulse to trigger *FF1* back to normal. *FF1* then triggers *FF2* to lower the grid of T_2 to −40 volts and release the first scan relay. When scan contact S_1 opens, *FF8A* will trigger *FF2* back, re-establishing the grid of T_2 at zero volts so that T_2 is able to carry the current of the new scan relay when contact S_2 opens. As *FF2* returns, it sends a pulse to both *FF3* and *FF9*. These are controlled by *FF11* in such fashion as to block the pulse to *FF9* allowing only the pulse to *FF3*. As *FF3* changes, the circuit has completed its first scan cycle and is ready to start printing the next. This occurs as before, with the same succession of the flip-flop operations leading to the closing and releasing of other scan relays as the scan system moves down the chain.

Eventually the end of the chain is reached, and the circuit takes a slightly different sequence to get back to the beginning. This occurs as *FF1* is triggered which opens T_2 to close the next scan relay. Since there are no more scan relays in operated units to the right the current flows to *FF10*. A pulse from *FF10* triggers both *FF1* and *FF11* back to their stand-by states. As before, *FF1* triggers *FF2*, which drops the grid of T_2 to −40 volts, and releases the one scan relay remaining closed. As before, a pulse from *FF8A* returns *FF2* to normal, which in turn sends pulses to *FF3* and *FF9*. Now a different situation occurs: *FF11* is back in its stand-by state as a result of the signal from the end of the chain. The pulse to *FF3* is therefore blocked and the pulse is passed to *FF9* triggering *FF9* and returning the grid of T_1 to zero. This completes the round and the circuit is back at its stand-by state. If operations have occurred in the earlier units while the later ones are being printed, the cycle will restart. If all operations have been cleared from the storage units the circuit waits in the stand-by state, for the next operation.

References

1. The Operations Recorder—A High-Speed Printing Annunciator, J. R. Leslie, G. M. Keyser. *AIEE Transactions*, vol. 75, pt. III, June 1956, pp. 501–04.

2. Anelex Printer, Leo Rosen. *Joint AIEE, IRE, ACM Conference on Electronic Computers*, Institute of Radio Engineers, New York, N. Y., Association for Computing Machinery, Pittsburgh, Pa., Dec. 10, 1952.

The Hands-Free Telephone

V. RODEK
NONMEMBER AIEE

AMONG THE UTILITIES which contribute to good standard of living, the telephone stands in prominence, playing a vital role in the activities of many people. The ease of communication provided by the telephone aids and promotes progress in every phase of human endeavor. Since its invention much work has been done to improve the telephone and to make it adaptable for new communication services. Recent advances in the electronics art, resulting in component improvement and miniaturization, provide powerful means for furthering the improvement of the telephone. One example of telephone adaptation for rendering a novel service is the development of a hands-free telephone, also known as a loudspeaking telephone.

The purpose of this paper is to outline the design considerations involved in the production of the instrument and to describe a telephone produced by Stromberg-Carlson Company, a Division of General Dynamics Corporation, and distributed to the independent telephone industry.

Function of Hands-Free Telephone

The principal function of this instrument is to enable one to receive and transmit conversation without having to use a handset. In typical use, a person would be seated a few feet away from the telephone and would carry on a conversation with the distant party in the same manner that he would if the party were in his presence. In this manner the user may write notes, refer to files, or perform any other duty which may be required during the course of conversation. In brief, during conversation the user has his hands free.

Hands-Free Telephone Components

In addition to a standard telephone unit including a handset, the hands-free telephone consists of a loudspeaker, a microphone or transmitter, two amplifiers, power supply, hybrid network, relay and associated circuitry. The loudspeaker, amplifiers, power supply, hybrid network, and circuitry are contained on one unit base. The transfer relay is contained within the standard telephone housing and microphone is external to the unit and is connected to the unit by means of a cable. Incoming voice signals are received from the loudspeaker driven by its associated amplifier and the outgoing voice signals are transmitted by means of the microphone and its associated amplifier. The hybrid network couples the telephone line to the transmitting and receiving amplifiers. The transfer relay functions as a control unit switching to handset or hands-free operation as required. The instrument is shown in Figs. 1 through 5.

Operation and Features

The design of the instrument provides features which make it easy to use. In originating a call, an ON switch button is depressed momentarily, the pilot lamp lights, the relay operates and the hands-free telephone is connected to the line. Dial tone or an operator is heard from the loudspeaker and one can now proceed to dial a number or instruct the operator by talking into the microphone. Conversation is carried on through the microphone and loudspeaker and at the end of the conversation the OFF switch button is depressed momentarily to disconnect the hands-free unit from the line. On answering a call the ON button is depressed momentarily and conversation may proceed after a warm-up period of approximately $2^{1}/_{2}$ seconds. During conversation the loudspeaker output may be adjusted to a suitable level by means of a volume control located at the front of the unit. If at any time it is required to suppress the transmission of outgoing speech or noise, the microphone cutoff button may be depressed and held as long as transmission is to be suppressed.

Provision is made for transferring from hands-free to handset operation. When the handset is lifted, the unit is automatically transferred to hand-set operation and the hands-free feature is made inoperative. To transfer a call from hand-set to hands-free operation the ON switch is depressed while the handset is being replaced on the cradle.

In using a loudspeaking telephone it is important to have as much separation as possible between the microphone and the

---

Paper 57-138, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 17, 1956; made available for printing November 26, 1956.

V. Rodek is with the Stromberg-Carlson Company, Rochester, N. Y.

Fig. 1 (left). Hands-free telephone in use as a conventional telephone

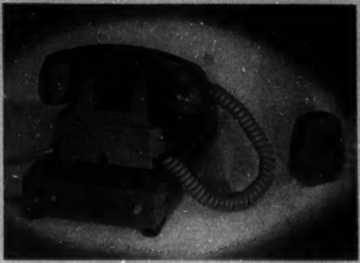


Fig. 3 (right). Rear view of hands-free telephone

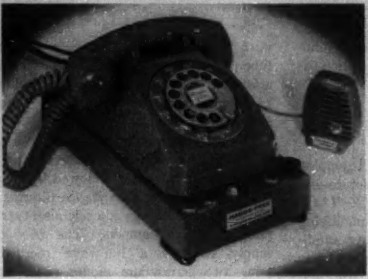


Fig. 2 (left). Front view of hands-free telephone

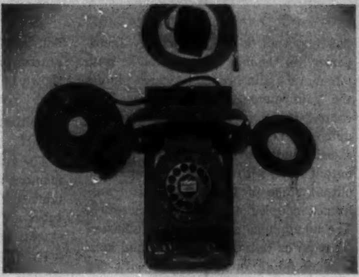


Fig. 4 (right). Top view of hands-free telephone

loudspeaker. This arrangement is necessary to minimize the feedback caused by the loudspeaker to microphone acoustical coupling which is inherent in all devices of this type. In a typical desk arrangement, the telephone may be placed on the left-hand side and the microphone on the right-hand side of the desk. The recommended distance between the microphone and the user is about 2 feet. When the location is quiet and the room or office has good acoustical properties the user may be several feet away from the microphone and carry on a conversation with ease.

To prevent the effects of mechanical coupling and to minimize acoustical coupling the microphone and loudspeaker are not mounted on a common structure. Either the microphone or the loudspeaker may be included in the main unit. In the hands-free telephone the microphone is used as a separate unit. This arrangement allows the microphone to be passed around as is sometimes advantageous during a conference call. Suitable microphone location for most convenient use and best performance is more easily attained when the microphone is separate from the telephone.

The housing enclosing amplifier sections and loudspeaker is 6$^1/_4$ inches wide, 12$^1/_2$ inches long and 1$^1/_2$ inches deep. The handset telephone is mounted on top of the housing and the loudspeaker grill is located behind the handset telephone. Line and power cords and the microphone cable are connected at the back end of the unit.

The color of the hands-free telephone is gray. This color is today's standard for office furniture and modern office machines. Gray will also harmonize with modern, colorful home furnishings.

Room Acoustics and Noise

The effects of room reverberation and noise are more prominent when using a loudspeaking telephone than when using a standard telephone. Room acoustics influence the degree of acoustical coupling between the microphone and the loudspeaker. The effect of acoustical coupling is "singing," a result of a feedback condition within the microphone, loudspeaker, and circuit loop. Rooms having good sound-absorbing properties allow more loudspeaker gain before singing occurs, thereby permitting higher loudspeaker volume levels.

A loudspeaking telephone in a reverberant room may cause the party at the distant end to hear his own voice as an echo. This reflection, transmitted by a loudspeaker telephone is particularly evident in reverberant rooms when the loudspeaker volume is at a high level.

Room noise has a tendency to be more interfering with a loudspeaking telephone than with a handset telephone. On receiving with a handset telephone, the receiver is placed against the ear and this helps to exclude interfering noise from the ear. With a loudspeaking telephone the listener is removed a few feet from the loudspeaker and the unattenuated room noise is received along with the transmitting speech.

Room acoustics and noise affect the performance of loudspeaking telephones. In installations where some care is taken to improve the room acoustics, as in homes and offices with drapes and carpeting, the performance of the hands-free telephone is very satisfactory.

The Amplifiers

The hands-free telephone includes two amplifiers, one for the microphone channel and one for the loudspeaker channel. Each amplifier is a resistance-coupled type. The microphone amplifier consists of three stages and has a maximum gain of 85 db (decibels). Its maximum

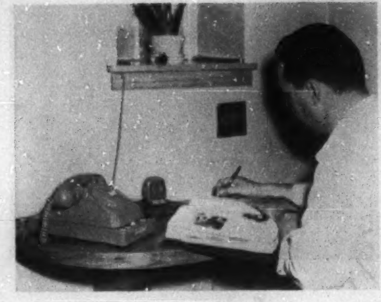


Fig. 5 (left). Hands-free telephone in use as a loudspeaking telephone

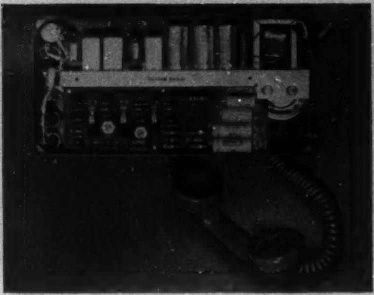


Fig. 6 (right). View of amplifier section with bottom cover removed

power output is 10 milliwatts. Incorporated in the microphone amplifier is an automatic gain control which limits the gain on high-input signals. This control functions to reduce the output difference between loud and soft speech signals and provides a higher average output level, with lower distortion. The output stage of the amplifier is transformer coupled to a hybrid network. The microphone channel, when adjusted for required output, has a noise and hum level of 55 db under output with the microphone short-circuited or disconnected.

A germanium diode supplies automatic gain control bias by rectifying the output signal of the microphone channel output stage. Distortion as a result of the action of the automatic gain control circuit is small because of its long time constant. The time constant of the filter network associated with the automatic gain control circuit is fairly long so that the bias will not follow any audio signal that the system is capable of producing. The signal level into the controlled stage is microphone level and is normally on a fairly linear portion of the tube transfer curve.

The loudspeaker amplifier consists of two stages. Its maximum gain is 35 db and it is capable of delivering 140 milliwatts into a 12-ohm loudspeaker. An input transformer couples the hybrid network to the input grid through a volume control. An output transformer matches the plate of the power tube to the loudspeaker. A signal of −10 to −15 dbm (decibels referred to one milliwatt) on the line will drive the loudspeaker to full output when the volume control is advanced to maximum.

Printed circuitry is employed in the assembly of amplifier sections. Use of printed circuitry makes possible a simplified assembly and a more uniform product, and allows greater degree of miniaturization. The amplifier section is shown in Fig. 6.

Amplifier Tubes

The amplifiers were designed to employ filament-type tubes of subminiature and miniature construction. *CK-512AX* tubes are used in the voltage-amplifying stages, and *3V4* tubes as power amplifiers. The tubes selected feature low current drain, long life, small size, and low microphonic level. One miniature type of power amplifier tube is used in the output stage of each amplifier.

Audio Transformers

The amplifier sections include three audio transformers. The loudspeaker matching transformer is designed to handle 250 milliwatts of audio power with a frequency response of 200 to 8,000 cps (cycles per second) ±1.5 db. The frequency response of the loudspeaker amplifier input transformer at one milliwatt of audio power is 200 to 8,000 cps ±1.5 db. The microphone channel output transformer secondary winding is bifilar wound to obtain an accurate impedance balance and optimum coupling between the two halves of the winding. This transformer is required to deliver 50 milliwatts of audio power with a frequency response of 200 to 8,000 cps ±1.5 db.

Power Supply

The hands-free telephone is designed to operate on a source of 117 volts a-c, 50 to 60 cps. The power transformer is connected directly to the power supply. The transformer has two secondaries; one secondary feeds a rectifier-filter network which supplies 120 volts d-c for tube plate potentials, and the other feeds a rectifier-filter network which supplies 24 volts d-c for the relay, pilot lamp, and tube filaments.

The tube plate voltage circuit consists of a half-wave rectifier, a surge resistor, and an input capacitor, followed by two stages of resistance-capacitance filtering. The half-wave rectifier is a selenium type rated at 50 milliampers. The 24-volt supply employs a full-wave rectifier using two germanium diodes rated at 150 milliamperes. The rectifiers feed an input capacitor from which is obtained current for the relay and pilot lamp. Two filter networks are taken from the input capacitor, each network containing two resistance-capacitance sections. The output voltage of these filters is 5 volts, and this voltage is used for tube filaments. Each tube is decoupled from the next one in its string by a large electrolytic capacitor. The decoupling insures low hum and minimum interaction between stages.

The power transformer floats on the line continuously. During the stand-by condition there is no filament current and the plate supply is idle. This condition would normally cause the $B+$ supply to rise to a value higher than that during operation and would momentarily place higher-than-normal plate voltage on the tubes when the set is turned on. This condition is prevented by placing a bleeder in the $B+$ filter circuit. The bleeder maintains the filter capacitors charged and prevents the $B+$ voltage from rising above its normal value. The resistor used to reduce the $B+$ during stand-by is shorted out by relay contacts when the set is turned on. The power which is consumed during stand-by is equal to approximately 5 watts, while during operation the power consumption is approximately 10.5 watts.

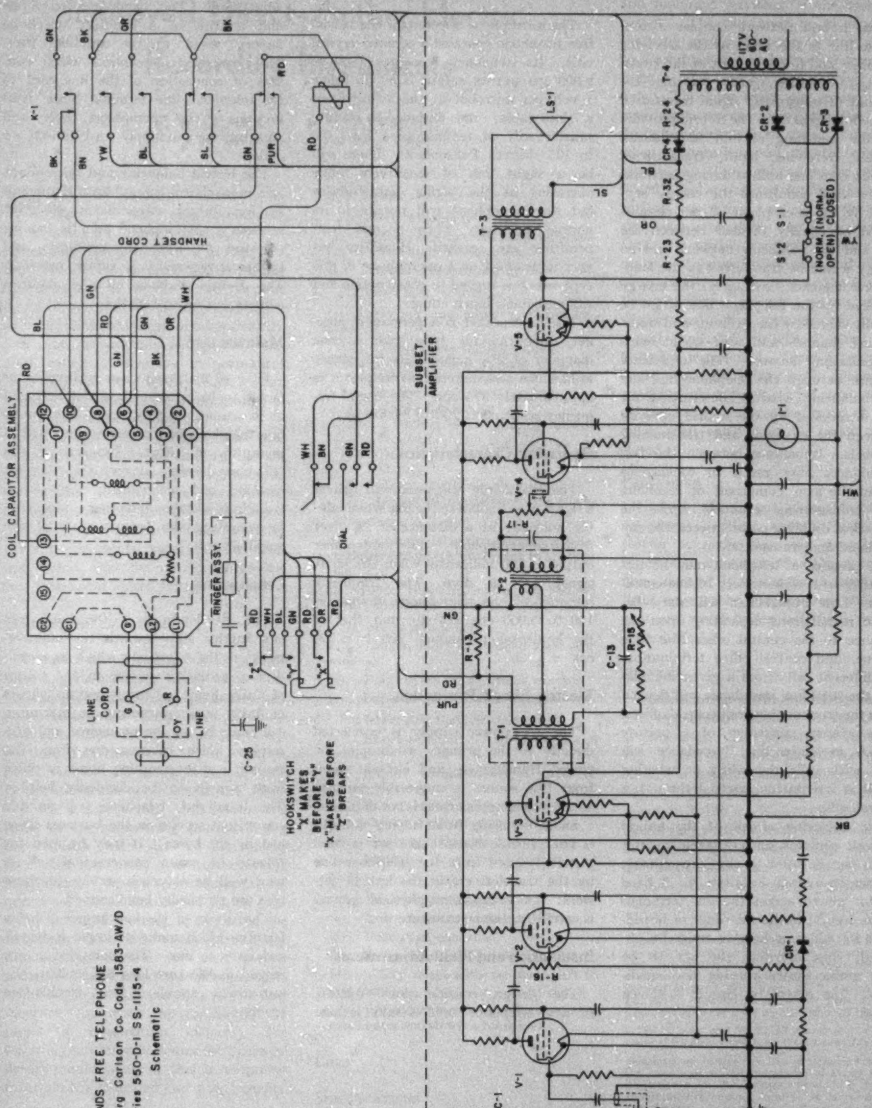

HANDS FREE TELEPHONE
Stromberg Carlson Co. Code 1583-AW/D
Series 550-D-1 SS-1115-4
Schematic
K-1
COIL CAPACITOR ASSEMBLY
HANDSET CORD
RINGER ASSY.
LINE CORD
LINE
C-25
HOOKSWITCH
"X" MAKES BEFORE "Y"
"X" MAKES BEFORE "Z" BREAKS
DIAL
SUBSET
AMPLIFIER
V-1
V-2
V-3
V-4
V-5
T-1
T-2
T-3
T-4
LS-1
MC-1
J-1
S-3 (NORM. OPEN)
S-2 (NORM. OPEN)
S-1 (NORM. CLOSED)
CR-1
CR-2
CR-3
CR-4
R-13
R-15
R-16
R-17
R-23
R-24
R-32
C-13
I-1
117V 60~ AC

Fig. 7. Schematic circuit of hands-free telephone

The Hybrid Network

The hybrid network couples a telephone line to the input of the receiving amplifier and to the output of the transmitting amplifier, and controls the amount of transmitter signal transferred to the loudspeaker. The hybrid provides for the matching of four impedances, namely, receiving input, transmitting output, and line and balancing network. Under ideal conditions the energy supplied by the microphone output channel would be equally divided between the line and the balancing network and no energy would be transferred to the loudspeaker channel. Similarly, the energy supplied by the telephone line would be equally shared by the receiving and transmitting channels and none would reach the balancing network. The impedance balance between the telephone line and the balancing network determines the loss introduced by the hybrid network between the receiving and transmitting channels. Unbalance between the two impedances may result in oscillations within the loop comprised of receiving and transmitting channels and the acoustical coupling path between the microphone and transmitter.

In practice a telephone may be terminated to a wide variety of line conditions. Line impedances will vary between installations depending upon the distance to the central office, the type of line, and central office termination. On different calls from a given installation the total line impedance will depend upon the trunking between central offices. In a system consisting of a private branch exchange line, impedance will vary with calls depending on whether a call is internal or external through a central office.

The balancing circuit of the hybrid network contains an adjustable resistor which can be used to obtain a suitable balance on a given installation. A fixed resistor placed across the line terminals of the hybrid limits the range of impedances for which a balance must be obtained, thus allowing the set to be more stable under varying line conditions. The schematic circuit is shown in Fig. 7.

Microphone and Loudspeaker

The microphone selected for the hands-free telephone contains a ceramic crystal unit. Its frequency response is 60 to 9,000 cps and its output is 50 db below 1 volt per microbar at the terminals of a 7-foot cable. It is designed to operate satisfactorily at temperatures from 0.0 to 135 degrees Fahrenheit. There will be a slight loss of sensitivity while operating at the higher temperatures but the microphone will restore to its normal sensitivity when normal temperatures are reached. Humidity has very little effect on a microphone of this type which is rugged in construction and will withstand much abuse.

The loudspeaker is a permanent-magnet, direct-radiator type with a cone diameter of $2^{3}/_{4}$ inches. Its impedance is 12 ohms and its resonant frequency is approximately 375 cps. The useful frequency response is 200 to 8,000 cps.

Operating Characteristics

The hands-free telephone will deliver a signal of −10 dbm to the line when talking normally at a distance of $2^{1}/_{2}$ feet from the microphone. The loudspeaker output is 100 milliwatts when the input signal is −10 dbm. The frequency response of the microphone channel is 100 to 5,000 cps ±3 db, and that of the loudspeaker channel, 200 to 5,000 cps ±5 db.

Electric Shock Prevention

The a-c power supply is connected directly to the primary winding of the power transformer, and electric shock from this source is impossible as long as the power transformer is not defective.

Shock resulting from ringing voltages is not possible because the set is completely isolated from the telephone line by the transformers in the hybrid network. In addition, a physical ground is carried to the instrument.

Installation and Field Adjustments

The design considerations relating to any telephone unit should include ease of installation and minimum field adjustment. The hands-free telephone unit consists of a hand-set telephone factory wired to the amplifier unit. This simplifies installation which consists of connection of the line cord to the telephone line terminal block, connection of the microphone cable, and plugging the a-c supply cord into an a-c outlet.

The hybrid balancing and microphone gain controls are accessible to the installer for further adjustments whenever necessary. Microphone gain is factory adjusted for optimum operation and further adjustment is rarely required. The hybrid balance usually requires adjustment on installation.

Maintenance

Ease of maintenance is an important factor to be considered in the design of an instrument of this type. The hands-free telephone can be maintained with a minimum of effort. Removal of the telephone housing exposes the relay for cleaning and adjustment, and access is made to all tubes and other components by removing the bottom plate of the amplifier unit.

Conclusions

The development of the hands-free is a further step towards the improvement of the telephone which now renders a necessary service in the course of daily living. During certain phases of daily work, answering or making a call may be an inconvenience and any service which releases one from the necessity of interrupting home or office work proves to be decidedly helpful. The hands-free telephone will provide convenient service in the business office and in the home. It may be used for private or group conversation and it may well be of much service to those who are physically handicapped.

The design of the instrument includes features which make it simple to install and easy to use. The components are rugged and feature long life, making the instrument capable of long, trouble-free service.

Improved Direct-Acting Electric Recording Instruments

E. W. CLARK
MEMBER AIEE

H. E. ALBRIGHT
NONMEMBER AIEE

WITH the electrical industry's increase in size and importance throughout the United States, a corresponding increase occurs in applications for electric direct-acting recording instruments. As power levels become larger, it becomes increasingly important to improve their control and application. Recording instruments possess the valuable feature of showing in permanent form the history of a given circuit. From this history, the future performance of the circuit can be predicted.

A thorough study revealed that contemporary recording instruments did not adequately fulfill present requirements and those predicted for the future. Specifications were formulated for a new line of recording instruments to meet these requirements, samples were built and tested, and the new recorder line was tooled and placed into production.

General Specifications

The over-all objective was to provide recording instruments which were more versatile, sturdier, more accurate, and more economical than any currently available. Before formulating the specifications for this recorder, a comprehensive review, started in 1949, was made to determine the applications and desirable features required. This review was made by the engineering and marketing departments with district instrument specialists and many users of recording instruments. Severe environmental requirements were established which generally included meeting or bettering the American Standard specifications for direct-acting electric recording instruments.[1] The detailed specifications covering the four main portions of the new recorder comprised approximately 11 pages of typewritten material. The following summary indicates the principal design requirements of this development to attain the indicated customer benefits.

1. For more accurate records:
   (*a*). Less affected by external influences.
   (*b*). Higher torque-to-weight ratio of moving system.
   (*c*). Less dimensional change in paper and improved mechanical timing drive.
   (*d*). Faster response time.
2. For more readable records:
   (*a*). Better color contrast between printing on chart and recording ink.
   (*b*). Superior blueprinting qualities.
3. For lower cost records:
   (*a*). Less ink and paper servicing because of
      (1). Three times the ink supply.
      (2). Inkwell sealed against evaporation and, entrance of foreign matter.
      (3). Visible ink supply.
      (4). No filling or cleaning of inkwell required.
      (5). One and one-half times the paper storage.
      (6). Paper supply clearly visible.
      (7) Longer running time spring clock.
4. For clearer records:
   (*a*). Ink supplied in three drying speeds; customer can match ink drying to application.
   (*b*) Faster response, mechanism follows circuit conditions more faithfully.
5. For better indication:
   (*a*). Larger, more easily read scale near window.
   (*b*). Point of recording nearer window.
   (*c*). Superior internal illumination.
6. For lower cost maintenance:
   (*a*). Shock-resistant spring-mounted jewels.
   (*b*). Higher torque-to-weight ratio of moving system.
   (*c*). More accurate and longer lasting clock drive with jeweled movement.
   (*d*). Nonbreakable pens.
   (*e*). Recorder easy to disassemble and repair.
7. For more versatile application:
   (*a*). Lower volt-ampere burdens.
   (*b*). Faster response time.
   (*c*). Twice the range of chart speeds.
   (*d*). Wider range zero set (zero left to zero right).
   (*e*). Wider temperature applications with case heater and slower drying inks.

Case

Specifications

Specifications require the housing for the instrument to have a pleasing appearance, to be weatherproof, and sturdily built, to contain the necessary resistors, and to have sufficient hand room for easy and foolproof servicing of the marking system. The case must be so arranged that the major components may easily be assembled into and removed from the front of the case by use of hand tools only. The case must also protect the chart-drive mechanisms and the pen-drive mechanisms from abuse, tampering when unattended, and adverse climatic conditions. Two applications for switchboard cases are needed for semiflush and surface mounting, and one portable-style case is required. In many applications the visibility of the pointer and scale is of prime importance and thus a large, clearly read scale near the front of the recorder is included as a "must" specification. For switchboard instruments, uniformly bright internal illumination is required. Since a recording instrument case is frequently opened for servicing ink or paper, or for note making on the record paper, the recorder case must open easily and be sealed when closed.

Procedure

Early in the layout phase of the development, styling experts were employed to make this new recorder as superior in appearance as the specifications insured it would be in operation. Some appearance samples were made of clay and wood before any final designs were made.

Results

Co-ordination of functional operations and appearance resulted in the design of two cases. The portable is shown in Fig. 1 and the switchboard for either surface or semiflush mounting is shown in Fig. 2. Case dimensions for both are shown in Table I. Weights of the instruments vary with the combination selected, averaging approximately 33 pounds. The switchboard case has no projecting knobs, hinges, or handles. Finish for

---

Paper 57-220, recommended by the AIEE Recording and Controlling Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted June 1, 1956; made available for printing December 5, 1956.

E. W. CLARK and H. E. ALBRIGHT are with the General Electric Company, West Lynn, Mass.

Development of this new line of recorders required the co-operation and knowledge of many people, particularly C. B. Huston, Steel Mill Section, Industrial Engineering; and R. B. Kerr, Appearance Design; E. R. Harrison, Electric Instrument Sales; and F. M. Learned, Electric Instrument Drafting; all of the Instrument Department, General Electric Company. Particular credit should be given to the users of recording instruments who gave of their time and effort during our user interviews. These were the Meter and Instrument personnel of many steel, industrial, and public utility companies.

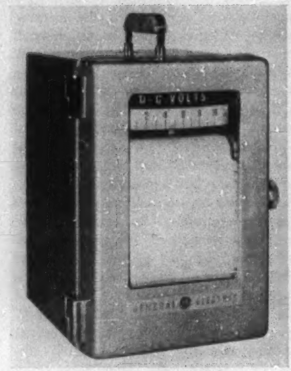


Fig. 1. Portable recorder

the switchboard case is a nonreflecting dull black. The portable case is finished in light-grey hammertone.

Paper-Chart Drive

SPECIFICATIONS

During the formulation of the specifications it was realized that a great many more chart speeds are required than are found in contemporary recorders, and that speeds must be easy to change. Servicing of the recorder was to be lessened by increasing the amount of paper which the chart carriage could store and by providing a suitable supply indicator to show the unused paper remaining on the supply roll. Actual values are shown in the results. The spring-clock chart-drive mechanisms must run satisfactorily for longer periods between servicings, must provide wider ranges of chart speeds, and at the same time must be much easier to disassemble for maintenance. Means were required for coupling together separate chart carriages for synchronization. All chart-drive mechanisms were to be completely interchangeable in all recorders.

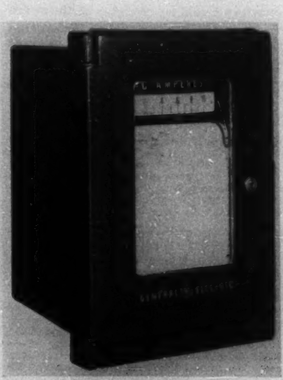


Fig. 2. Switchboard recorder

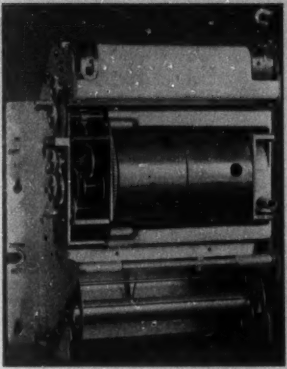


Fig. 3. Chart-drive mechanism with front plate removed to show spring-clock driving mechanism

PROCEDURE

Investigation of prior art, calculations, layouts, and tests demonstrated that five styles of chart-drive mechanisms were required to fulfill satisfactorily the various requirements. These styles include low-speed electric drive, high-speed electric drive, multispeed electric drive with either internal or external switching, and a completely new spring-clock driving mechanism to provide both increased running time and increased torque.

RESULTS

The chart carriages are designed to store and reroll 150 feet (50 feet more than contemporary designs) of 6-inch-wide paper and provide 28 different speeds ranging from 1/4 inch per hour to 120 inches per minute. High-speed changes from inches per hour to inches per minute are made by changing lever position on the chart carriage. A small pointer located on the lower left of the chart carriage indicates the amount of paper remaining on the supply bobbin. Four styles of chart carriage employ synchronous electric motors for the drive power while the fifth uses a completely new spring-clock mechanism, Fig. 3.

A chart-drive mechanism, since it is a rather complicated collection of gearings and motors, requires maintenance similar to any mechanical assembly. No adjustment of shaft end play or timing-drum end play is required upon reassembly of this chart-drive mechanism.

Marking Systems

INTRODUCTION

Many variations of marking systems which use ink have been employed in recorders over the years. Although there have been many physical shapes and sizes, there has always been one common feature of these systems, namely, the requirement of transferring ink from a bottle to the inkwell and pen by the means of an eyedropper type of syringe and tube. This is an operation which has many drawbacks. Ink is quite often spilled on either the operator or the recorder. In addition, the ink in the supply bottle ages and thickens. Inkwells allow evaporation and the entrance of foreign matter. Pens become clogged with old ink or foreign matter and cease to make a record. Other types of marking systems alleviate to some extent some of the hazards of an inking system, but possess other shortcomings such as discontinuous records, external power requirements, or limited pen-to-paper speed ranges. Most of the troubles associated with the use of ink in a recorder were associated with the transferring of the ink from the supply bottle to the inkwell, the evaporation of the ink in the inkwell, and the contamination of the ink in the inkwell or supply bottle.

Table I. Over-all Dimensions

Case Type	Height, Inches	Width, Inches	Depth, Inches
Semiflush mounting switchboard	$16\,^{3}/_{16}$	$10\,^{7}/_{8}$	$12\,^{5}/_{8}$
Surface mounting switchboard	$16\,^{3}/_{16}$	$10\,^{7}/_{8}$	$12\,^{5}/_{8}$
Portable	17 [illegible]/32	$10\,^{15}/_{32}$	13

Table II. Ink Drying Speeds

Color	Fast Drying (Fast Chart Speeds), 32 F to 95 F	Medium Drying (Medium Chart Speeds), 20 F to 104 F	Slow Drying (Slow Chart Speeds), −5 F to 120 F
Red	G1	G2	G3
Black	G4	G5	G6
Green	G7	G8	G9
Brown	G10	G11	G12

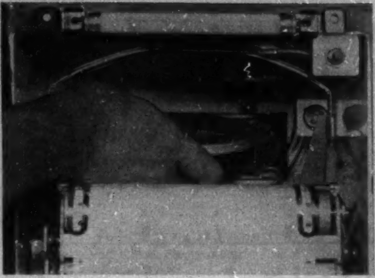


Fig. 4. Inkwell installation or removal

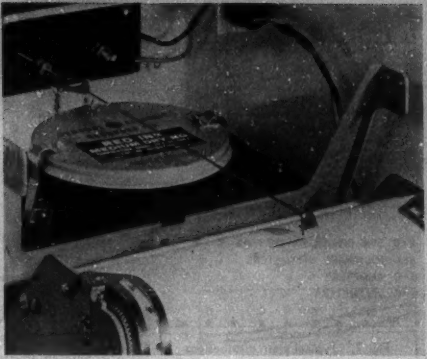


Fig. 5 (right). Complete marking system in operating condition

SPECIFICATIONS

A new type of marking system must require less time and skill to service. It must be economical, and nontoxic, must possess superior aging qualities, and must be more certain of operation over wider ranges of operating conditions. A clear definite record is required over the entire range of pen-to-paper speeds of 0.004 to 600 inches per minute.

PROCEDURE

Over a year of investigation revealed that the advantages of ink outweighed the advantages and disadvantages of other available marking means. The procedure was then clearly defined as one of eliminating all the major disadvantages of ink in present-day inking systems and minimizing the remaining ones.

RESULTS

An inkwell was developed which is used as the supply bottle for the ink for this recorder. It is strong, transparent, and low enough in cost to allow the inkwell to be thrown away on depletion of the ink supply. The opening in the inkwell is only 5/32 inch in diameter and greatly reduces ink evaporation and entrance of dirt into the ink-supply system. The usual type of inkwell has a curved opening approximately 1/4 inch by 1 inch or larger. The throwaway combination inkwell and bottle contains 3 ounces of ink which is approximately three times as much as can be held in the conventional inkwell. The clear plastic makes it easy to see the amount of ink remaining and the throwaway feature eliminates the nuisance and expense of cleaning and refilling inkwells with the associated possibility of ink spillage. The small entrance hole makes possible the transportation of recorders in normal operating position without fear of ink spillage. Superior blueprinting of the records and good optical contrast is made possible by choice of proper ink dyes. Special inks were developed for extremes in operating conditions, and different colors are available as indicated in Table II. The special precious-metal-alloy pen tip was designed for extremely long life and negligible wear as compared to other types of material often used for this purpose. Fig. 4 shows the inkwell being installed and Fig. 5 shows it and the pen in normal operating position.

Measuring Mechanisms (Pen-Drive Mechanism)

GENERAL SPECIFICATIONS

All mechanisms must be designed for a 1% accuracy class rating. These mechanisms must be sturdy, and so arranged as to provide electrical insulation protection for the operating personnel when changing pen, ink, or chart paper. The mechanisms must be protected from ink spillage and entrance of foreign matter when the case is open. A zero set must provide at least half-scale range, and preferably full-scale range, to permit adjustment of the electrical zero. Response time for the mechanisms must be in the half-second class. Instrument burden should be as low as possible, consistent with these performance levels. Easily adjustable magnetic damping is required. The insulation level of the measuring circuit must be in the 5,000-volt high-potential class.

GENERAL PROCEDURE

Careful measurements were made to determine the actual friction between the chart paper and pen under normal operating conditions. From these measurements and from past history of recording instruments the minimum full-scale torque value could be determined for an allowable pen-friction error. Over half a century of recording instrument experience has demonstrated that pivot friction in well-built vertical-shaft recorders is negligibly small when compared to the pen friction.

$$T = \frac{T_p}{e} \times 100$$

where

T = full-scale torque
T_p = pen-friction torque
e = allowable pen-friction error, per cent of full scale

A second factor which affects the required torque level of the recorder is the response-time requirement. Response

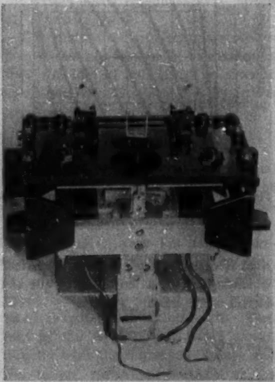


Fig. 6. A-c ammeter mechanism

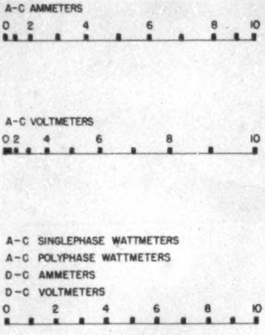

A-C AMMETERS
0 2 4 6 8 10
A-C VOLTMETERS
0 2 4 6 8 10
A-C SINGLEPHASE WATTMETERS
A-C POLYPHASE WATTMETERS
D-C AMMETERS
D-C VOLTMETERS
0 2 4 6 8 10

Fig. 7. Typical scale distributions

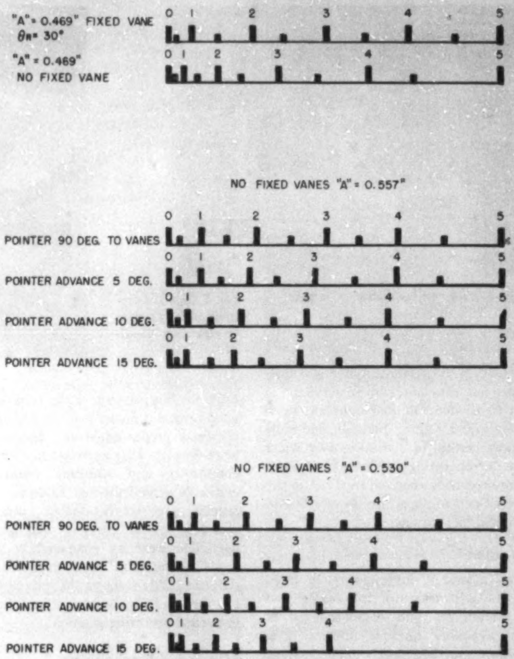

"A" = 0.469" FIXED VANE
θR = 30°
"A" = 0.469"
NO FIXED VANE
NO FIXED VANES "A" = 0.557"
POINTER 90 DEG. TO VANES
POINTER ADVANCE 5 DEG.
POINTER ADVANCE 10 DEG.
POINTER ADVANCE 15 DEG.
NO FIXED VANES "A" = 0.530"
POINTER 90 DEG. TO VANES
POINTER ADVANCE 5 DEG.
POINTER ADVANCE 10 DEG.
POINTER ADVANCE 15 DEG.
0 1 2 3 4 5

Fig. 10. A-c ammeter scale distributions for various combinations of vanes and poles

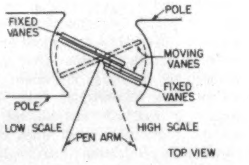

POLE
FIXED VANES
MOVING VANES
FIXED VANES
POLE
LOW SCALE
PEN ARM
HIGH SCALE
TOP VIEW

Fig. 8. A-c ammeter vane arrangement

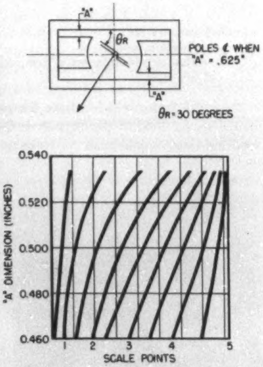


Fig. 9. Effect of pole shift on scale distribution of a-c ammeter mechanism

time is dependent upon the free period[2] which in turn is related to the torque and inertia of the moving system.

$$T_o+2\pi\sqrt{\frac{P}{t}}$$

where

T_o = time of an undamped vibration of movable system

P = moment of inertia of entire moving system

t = torsion constant of suspension, restoring moment for unit angular deflection

The inertia varies with the mass and geometry of the moving system and is minimized by choice of material, positioned as close to the axis of rotation as is compatible with functional requirements. Torque levels may be increased for any style of mechanism, but only by an increase in power input (volt-ampere burden) or mechanism size, weight, and inertia. For this class of instrument, the response time is approximately 0.5 to 0.75 of the free period.

GENERAL RESULTS

Highly efficient ironclad mechanisms for a-c and d-c measurements were required to accomplish the desired results. A-c mechanisms are required for the measurement of amperes, volts, singlephase watts, polyphase watts, reactive volt-amperes, and frequency. D-c mechanisms are required for the measurement of milliamperes, amperes, and volts. The mechanisms are designed to utilize common parts to the utmost. They are electrically insulated from operating personnel and are protected from ink spillage and foreign matter (see Fig. 6). All field structures consist of

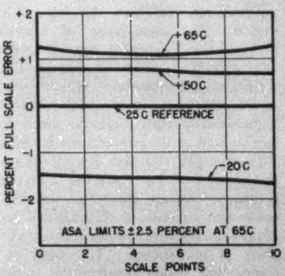


Fig. 11. Extreme temperature influence on a-c ammeter mechanism

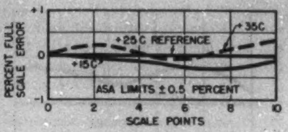


Fig. 12. External temperature influence on a-c ammeter mechanism

stacked laminations for the magnetic return path and either electromagnets or Alnico V permanent magnets for the pole pieces. All mechanisms use wound armature coils with the exception of the ammeter, which makes use of an iron-vane moving system. Torque levels were established to limit pen-friction error to 1/4 of 1% of full scale and to provide the desired approximate half-second response time for the various mechanisms with their different inertias. Variations in torque are between 100 millimeter-grams full scale for the a-c ammeter and 159 for the single-phase wattmeter. In addition to the standardization of parts in the mechanisms, it is highly desirable that scale distributions be made the same for as many different mechanisms as possible. Only three scale distributions are required for all standard measurements. This standardization helps both the user and manufacturer by requiring fewer charts in inventory. Typical scale distributions are shown in Fig. 7. A magnetic shield is employed with the a-c mechanisms to reduce the effect of external magnetic fields to values considerably below those required to meet ASA specifications.

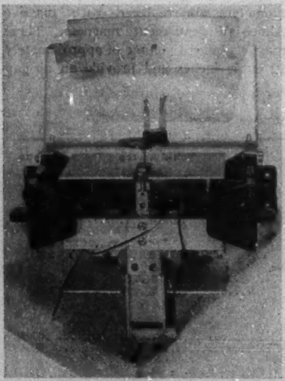


Fig. 13. A-c voltmeter mechanism

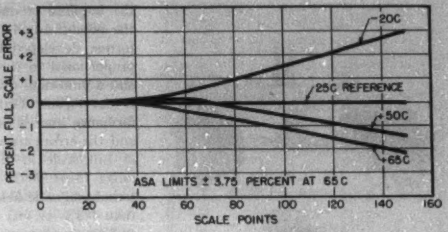


Fig. 14. Extreme temperature influence on a-c voltmeter mechanism

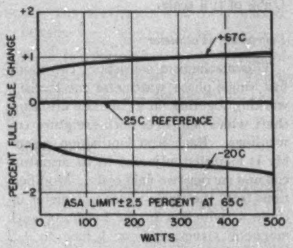


Fig. 15. Extreme temperature influence on a-c single-phase wattmeter

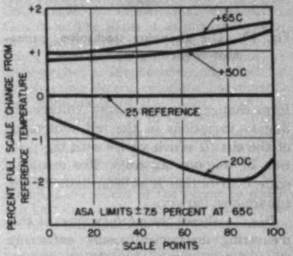


Fig. 16. Extreme temperature influence on a-c polyphase wattmeter

MECHANISMS FOR A-C MEASUREMENTS

A-C Ammeter Mechanism (*Fig. 6*)

The basic torque-producing parts of this unique mechanism are repulsion forces between fixed vanes and moving vanes and attraction forces between moving vanes and eccentric pole pieces. The flux which actuates the vanes is obtained from the external iron circuit employing two field coils connected in series or in parallel. The fields from these two coils are aiding and the coils are mounted one on each pole piece. The moving vanes consist of straight laminations through the armature shaft. The fixed vanes consist of offset laminations attached to the element frame. The two sets of vanes are so arranged, see Fig. 8, that the moving vanes are quite close to the fixed vanes at zero position; and as the moving vanes rotate up scale, they move away from the fixed vanes. The close coupling between the fixed and moving vanes at the low-scale end produces sufficient torque to expand the distribution over the lower section of the scale. The distribution of the upper portion of the scale is primarily controlled by the amount of the offset of the pole pieces, Fig. 9. With no fixed vane, the scale distribution is quite constricted at the low-scale end as indicated in Fig. 10. If no pole offset is employed, approximately only one-half scale deflection can be obtained. A combination of offset pole pieces and repulsion vanes results in a scale distribution for this unique mechanism which is superior to that obtained with an ordinary electrodynamometer. This can be seen by a comparison of the scale distributions for the a-c ammeter and a-c voltmeter (ordinary electrodynamometer) which are shown in Fig. 7.

The effect of external temperature influences on this measuring mechanism is shown in Fig. 11 and 12. The effect of an external magnetic field of 5 oersteds is less than 1/2%.

A-C Voltmeter Mechanism (*Fig. 13*)

This mechanism is an electrodynamometer of the ironclad variety. The windings on each pole piece are connected in series with the armature or moving coil, and a very low temperature-coefficient series resistance. The moving coil rotates in a uniform air gap between the pole pieces and an iron core, which is centered in the armature. Current in the series circuit varies almost directly with the impressed voltage and the torque varies substantially as the square of the current flowing in the series circuit. Deviations from the uni-

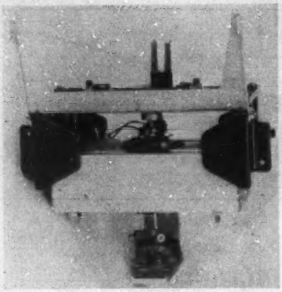


Fig. 17. D-c measuring mechanism, permanent-magnet moving-coil type

form change in current with voltage are due to variations in the self-inductance of the circuit which varies with the position of the moving coil. The resulting scale distribution is substantially of the square-law variety.

Fig. 14 shows the performance of this measuring mechanism under extremely adverse environmental conditions of temperature. Over the ASA temperature range of 15 to 35 C (degrees centigrade) errors did not exceed 0.4% of full scale. The effect of an external magnetic field of 5 oersteds is less than 0.8% of full scale. Condenser compensation is employed to reduce the frequency error to less than ±0.4% of full scale over the range of 25 to 70 cycles for a normal 60-cycle instrument.

Single-Phase Wattmeters

This mechanism is the same as the voltmeter except for the coils. The field coils are now connected in the current circuit; and the armature, in series with a resistance, is connected in the potential circuit. This system measures the in-phase product of the currents in the potential and current circuits. The torque developed by the mechanism is proportional to the watts in the circuit and a uniformly divided wattmeter scale is developed. Fig. 15 shows the performance over wide temperature ranges and the errors are not in excess of 0.3% of full scale in the ASA temperature range 15 to 35 C. The effect of an external magnetic field of 5 oersteds is less than 0.8% of full scale. Over the 50- to 70-cycle frequency range the deviation from 60-cycle indication is less than 1/4% of full scale.

Polyphase Wattmeter

This mechanism consists of two modified single-phase wattmeter mechanisms working together on a common armature shaft which supports both armature coil windings. Each iron lamination assembly is mounted adjacent to its armature coil and carries two field coils. Modifications from the single-phase wattmeter include a reduction in the height of each magnetic structure from 1 inch to 1/2 inch with a corresponding reduction in height of its armature coil. Each separate assembly develops approximately one half of the torque of the single-phase wattmeter, but the two assemblies develop a torque comparable to that of the single-phase wattmeter or a-c voltmeter. A magnetic shield is provided between the two mechanisms to reduce the interelement interference to less than 0.3% of full scale.

Fig. 16 shows the temperature effect over wide temperature ranges and the errors do not exceed 1/4% over the 15 to 35 C ASA temperature range. The effect of an external magnetic field of 5 oersteds is less than 1%. Over the 50- to 70-cycle frequency range the deviation from the 60-cycle calibration does not exceed 0.2% of full scale.

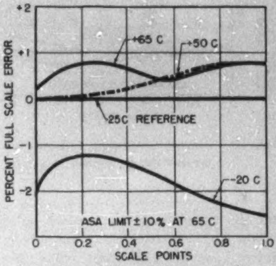


Fig. 19. Extreme temperature influence on d-c 1-milliampere mechanism

Varmeters

The measurement of reactive volt-amperes is more accurate and has generally replaced the measurement of power factor. This type of measurement is made with regular wattmeters plus external phase-shift networks common to the art.

Frequency Meter

This recording instrument consists of a zero-center signal phase watt-meter measuring mechanism with a suitable armature and field coil combined with an external frequency-sensitive impedor. The instrument employs new and unique circuitry, and a technical paper is now being written to describe it.

MECHANISM FOR D-C MEASUREMENTS (FIG. 17)

The d-c measuring mechanism is quite similar to the a-c mechanisms except that the lamination stack is 2 inches high and the electromagnets of the electrodynamometers are replaced with massive Alnico V permanent magnets. These magnets have a volume of approximately 12 cubic inches and provide an air-gap

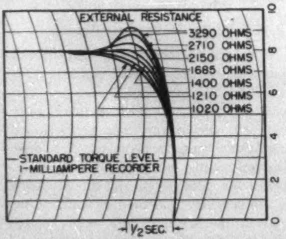


Fig. 18. Response curves for d-c measuring mechanism rated 1-milliampere full scale with internal damping omitted

Table III. Typical Instrument Losses

Instruments	Volt-Amperes	Watts
A-C:		
Ammeter 0–5 scale (at 5 amperes)	9.2	1.4
Voltmeter 0–150 scale (at 120 volts)	8.5	8.5
Single-phase wattmeter 120 volts, 5 amperes		
Potential circuit (at 120 volts)	7.8	7.8
Current circuit (at 5 amperes)	6.8	2.2
Polyphase wattmeter 120 volts, 5 amperes, 3 wire, 3 phase		
Potential circuit (at 120 volts)	5.8	5.8
Current circuit (at 5 amperes)	4.8	1.6
D-C:		
Ammeters		
0–1 milliampere (at full scale)		.0003
0–100 milliamperes (at full scale)		.0005
Over 20 amperes require external shunt, either 50 or 100 millivolts full scale		
Voltmeters		
Zero left or zero center, 200 ohms per volt.		

flux density of approximately 5,000 gauss. This high flux density has made it possible both to decrease the response time and lower the burden of the d-c measuring mechanism.

The 1-milliampere full-scale recording instrument has a response time of 0.6 second and a full-scale burden of 0.3 milliwatt which compares with the 1.2 to 1.5 milliwatts previously required. This low burden may be of major significance in certain applications such as the combination of the d-c measuring mechanism with electronic devices for extension of measurement.

As a result of the high flux density of this d-c measuring mechanism, considerable damping is obtained from the circuit unless the resistance of this circuit is greater than certain minimum values. In the case of the 1-milliampere full-scale-rated d-c mechanism, the minimum circuit resistance to give critical damping (with built-in damping adjusted to zero) is 1,250 ohms. Resistance less than this value will cause the recorder to be overdamped and the response time increased. Resistance above 15,000 ohms reduces the circuit damping to a negligible amount, and the built-in damping obtained by the use of damping magnet and damping vane is required. To allow a customer to adjust the damping of his d-c mechanism to suit his circuit conditions, the damping magnet is easily adjustable from zero to overdamped conditions. Fig. 18 shows a family of response-time curves for a standard 1-milliampere recorder.

In those cases where even faster response time is required and where additional burdens above 0.3 milliwatt can be tolerated, special high-speed d-c measuring mechanisms can be supplied. For example, if 5-milliampere direct-current into 300 ohms is permissible, then a response time of approximately 0.3 seconds can be obtained.

The d-c mechanism performance over wide temperature limits is indicated in Fig. 19 and errors do not exceed 0.2% full scale over the ASA temperature range 15 to 35 C. Effect of an external field of 5 oersteds could not be detected, and a maximum error of 0.3% of full scale was found when the field was increased to 20 oersteds.

D-c recorders are self-contained in ratings from 500 microamperes through 20 amperes. Ratings above 20 amperes require external shunts with output of 50 millivolts or above.

Voltmeters are designed for operation at 200 ohms per volt (5-milliampere armature) and are self-contained up to and including 750 volts. Low-burden d-c voltmeters can be made with resistance up to 2,000 ohms per volt.

SCALE DISTRIBUTIONS AND INSTRUMENT LOSSES

Scale distributions for typical types of recorders are shown in Fig. 7. Typical instrument loss data are shown in Table III.

Conclusions

Customers, salesmen, and application engineers were queried to determine the good and bad features of existing direct-acting electric recording instruments, and the desirable new features required for future applications. Replies from questionnaires and data obtained in interviews were compiled and specifications were formulated to cover a recorder line which would satisfy the requirements. The recorder was separated into its major components, that is, the case, the marking means, the chart carriage, and the measuring mechanisms. Designs were formulated and sample instruments were developed and tested for each of these major components to determine their compliance with the specifications. Results were continually reviewed with interested parties in marketing, manufacturing, and engineering fields. This co-ordinated program resulted in a new line of direct-acting electric recorders which have more desirable features for the ultimate user than do previously available types. These features include more accurate records, more readable records, lower cost records, clearer records, better indication, cheaper maintenance, more versatile application.

References

1. DIRECT-ACTING ELECTRICAL RECORDING INSTRUMENTS, SWITCHBOARD AND PORTABLE TYPES. *ASA Standard C-39.2-1953*, American Standards Association, New York, N. Y., 1953.

2. FUNCTIONAL ANALYSIS OF MEASUREMENTS, I. F. Kinnard. *AIEE Trransactions* (*Electrical Engineering*), vol. 65, 1946, pp. 987–95.

3. ELECTRICAL MEASUREMENTS (book), F. A. Laws. McGraw-Hill Book Company, Inc., New York, N. Y., 1938, p. 17.

◆

Discussion

K. M. Mitchell (The Hydro-Electric Power Commission of Ontario, Toronto, Ont., Canada): The paper does not make clear the extent to which this new line of recorders meets either the specification formulated as an objective, or the American Standard for direct-acting electric recording instruments (see reference 1 of the paper). Both of these are referred to in the paper. From the user's or consumer's point of view, *ASA C39.2-1953* can serve as a very useful basis for a performance appraisal of recorders of different manufacture. For this reason, a statement of the extent to which this new line of recorders meets the requirements of this standard would be of considerable value. Can it be stated that in all respects this new line of recorders meets the requirements of *ASA C39.2-1953?*

E. W. Clark and **H. E. Albright**: As the development of this new line of recorders was progressed and reviewed with interested parties, no changes in the major objectives were found necessary. A few minor modifications in specifications were made to improve over-all performance as typified by improved ratios for chart speeds, and a reduction in the allowable losses for the d-c milliammeters. A revision of this paper eliminated the statement that the new line of recorders meets all of the requirements of *ASA C-39.2-1953* except for one very minor deviation. This is in the frequency meter where the losses exceed the specification limit by a trivial amount under some operating conditions. Conversely, performance of the recorders is decidedly superior to the requirements in many respects as typified by scale length, response time, and extreme temperature performance. As a matter of interest the two most difficult parts of the specification to meet were our requirement of storing and rerolling 150 feet of chart paper, and the *ASA C-39.2-1953* requirements 6.2.1 on momentary overload where the pen has a great desire to part company with the pen stirrup.

The BTU Computer: An Advanced Data System for Industry

FRED W. HANNULA
NONMEMBER AIEE

AN industrial process is usually controlled by individually controlling all of its significant variables such as temperatures, pressures, and flows. The intelligence which establishes an optimum operation of the process must then be defined in terms of these variables.

So controlled, the process obviously requires the faithful performance of each unit. One variable out of control will upset the whole, for automatic readjustment of the other controllers is impossible. As a result, there is an increasing tendency to interlock several controllers to cause a mutual readjustment to meet changing needs. However, equivalent results can be often accomplished by manipulating measurements mathematically, in many cases reducing the multiplicity of controllers.

Often in a process, the parameter, which best describes its operation, cannot be measured directly but can be defined mathematically in terms of two or more directly measurable factors. At times this relationship is obscure and complicated; at other times it is exceedingly simple.

Increasingly, these key parameters are determined and special-purpose computers built to measure their values. The human operator is then supplied with more accurate and reliable information much more quickly than he could derive it from the "raw" data. Furthermore, automatic control can then be designed to produce a desired result rather than simply holding a group of variables within bounds.

Btu Computer Applications

The Btu computer is a special-purpose system with a remarkably wide field of application. Its function can be best described by relating it to one example, typical of many.

Fig. 1 shows a conventional heat exchanger. This may be part of a refrigeration system, a condenser, a chemical reaction vessel or an atomic reactor. The common condition among these is that there is an exchange of heat between the two materials. For discussion purposes, the through line is shown containing a cooling fluid which gains heat. The Btu computer is, of course, equally applicable when conditions are reversed, i.e., when heat is removed from this fluid by the material inside the exchanger.

In all of these devices, the rate at which heat is gained by the flowing medium is a measure of the level at which the unit is operating. The summation, or integration, of this rate over a period of time is a measure of the work done by the unit. An accurate measurement of the rate is often an invaluable means of control. An indication of the accumulation may be required either for control or for economic accounting when the work done is being sold, as in refrigeration units.

Btu Computer Function

The Btu computer is then an instrumentation system designed to measure the rate at which the cooling fluid gains heat and to integrate it with respect to time. Its function is most easily understood by expressing the problem in simple mathematical terms.

The fluid with a thermal capacity of C, Btu per pound per degree Fahrenheit (F), is flowing through the heating area at the rate of F pounds per minute. Each pound of the fluid enters the area at a temperature T_1, and, being obviously heated, leaves at a higher temperature T_2, gaining $C(T_2 - T_1)$ Btu. However, F such pounds per minute of fluid are flowing; therefore, the number of Btu's per minute, P, of heat absorbed is thus

$$P = FC(T_2 - T_1) \tag{1}$$

It should be noted that in equation 1 the units of P are in Btu per minute, a rate of heat transfer. This is a power unit and can be converted to other conventional power units, i.e., watts or horsepower, by simple multiplying factors.

By integrating equation 1 with respect to time, the total number of Btu absorbed during any given time interval can be determined.

$$W = \int P dt = C \int F(T_2 - T_1) dt \tag{2}$$

W is in Btu, a work unit, and can be converted to other work units, watthours or horsepower hours, for example, again by appropriate multipliers. The task of Btu computer is then to solve equations 1 and 2 instantaneously and continuously as the process progresses, i.e., in real time.

The terms in which the answer is presented must, of course, satisfy several demands. It must be capable of being recorded automatically. Means must be available for integrating the answer with respect to time (to obtain a solution to equation 2). Lastly, it will probably be necessary to operate a controller from this solution. An angular shaft position seems to satisfy all of these requirements.

Basic Design

In equation 1 there are two unknowns, the flow F and the temperature difference $(T_2 - T_1)$. The thermal capacity C is assumed constant.

The Btu computer first measures each of these variables and represents their values in terms of analog signals. These are manipulated to produce the required answers.

The establishing of the types of analogs to be used is perhaps the most important single decision required. On it rests, not only the basic computer design and the choice of measurement techniques, but the accuracy and the over-all cost of the system. An ingenious choice can pay tremendous dividends in simplification and reliability. In this computer, the ultimate selection is based on two unique electrical characteristics.

The first of these is the multiplying action of a conventional Wheatstone bridge. The second is the exponential characteristic of resistance versus temperature achieved in a nickel resistance bulb with selected series and shunt resistors.

Considering the bridge first, it is schematically shown in Fig. 2. Two conditions, which are easily fulfilled in practice, are assumed. First, the bridge is fed by a voltage E with a negligible source impedance in terms of the bridge impedance. Secondly, the output voltage E_0 is measured by a device whose impedance is very high compared with the bridge

---

Paper 57-222, recommended by the AIEE Recording and Controlling Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted July 10, 1956; made available for printing December 5, 1956.

FRED W. HANNULA was formerly with The Foxboro Company, Foxboro, Mass., and is now with the Applied Science Corporation of Princeton, Princeton, N. J.

The author wishes to express his appreciation to W. H. Howe of The Foxboro Company, Foxboro, Mass., for his invaluable suggestions and guidance in the preparation of this paper.

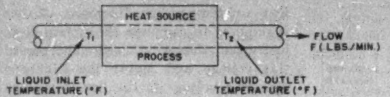

HEAT SOURCE
T_1
T_2
PROCESS
FLOW
F (LBS./MIN.)
LIQUID INLET TEMPERATURE (°F)
LIQUID OUTLET TEMPERATURE (°F)

Fig. 1. Typical heat exchanger

impedance. Thus, no significant loading effects are involved.

The upper two arms of the bridge contain two identical precision resistors R. Both of the lower arms are variable resistances (thermometer bulbs). The output voltage of this bridge can then be expressed as

$$E_0=\frac{E}{2}\left(\frac{R_2-R_1}{R_2+R_1}\right) \tag{3}$$

If the voltage E is made proportional to flow, and the resistive term to temperature difference, then equation 3 provides a simple means for solving equation 1.

Equation 3 makes no limitations on whether the voltage is direct or alternating. A 1,000-cps (cycles per second) voltage was chosen for several reasons.

Electronic equipment operating at audio-frequencies can be made simple, economical, and reliable in operation. Interference problems are minimized, and troublesome slide wires which are an inherent part of d-c computers can be eliminated.

Flow-to-Voltage Conversion

The first task is then to produce a voltage E proportional to flow F. This is a flow-to-voltage transducer.

A typical device is shown in Fig. 3. The flow is measured by an orifice which produces a differential pressure proportional to the square of flow. This differential pressure is very accurately converted to a shaft-angle position by means of a conventional mercury manometer.

To accomplish the necessary square-rooting action, this shaft angle is linked to a large accurately machined cam, so machined that its cam-follower angular position is directly proportional to flow. The follower position is then converted to a proportional 1,000-cps voltage (E) by the transducer transmitter, a linear differential transformer. The cam-follower movement is the last mechanical motion in the system with the exception of the pen which will ultimately record the Btu rate.

This voltage is then applied to the computing bridge (Fig. 2) as its input voltage E. The assumption is made that the density of the flowing fluid remains constant, therefore the rate of flow measured by the manometer is equivalent to the mass rate of flow prescribed by equation 1.

Other transducers may be used. A magnetic flow meter, with an output voltage inherently linear with flow, is ideal, because it eliminates the need for the orifice, the square-root extraction, and it operates over extreme flow rangeabilities.

Temperature-to-Resistance Conversion

In equation 3, the second requirement of equivalence is that the second term be directly proportional to temperature difference; on first inspection, a rather obscure task.

This term contains the two resistances R_1 and R_2. The resistance of thermometer bulbs varies with temperature. The resistances R_1 and R_2 can then represent bulbs at the inlet and outlet of the heat exchanger. Mathematically, the demand placed on the relationship of the temperatures and the bulb resistances can be represented by the following equation, where K is a proportionality constant, and x is the ratio of R_2 to R_1.

$$(T_2-T_1)=K\left(\frac{R_2-R_1}{R_2+R_1}\right)=K\left(\frac{x-1}{x+1}\right) \tag{4}$$

Here, the electrical characteristics of the adjusted nickel resistance bulb become important. In this type of bulb, the relationship between resistance and temperature can be expressed as

$$r=B\epsilon^{kT} \tag{5}$$

If then, in the bridge of Fig. 2, a bulb of this type is used to measure the inlet and outlet temperatures of the heat exchanger, the obtained ratio of the two resistances, x, is

$$x=\frac{R_2}{R_1}=\frac{B\epsilon^{kT_2}}{B\epsilon^{kT_1}}=\epsilon^{k(T_2-T_1)} \tag{6}$$

taking the logarithm of both sides of equation 6 the following relationship is obtained.

$$T_2-T_1=\frac{1}{k}\ln x \tag{7}$$

Using the series expansion of $\ln x$ in equation 7 gives

$$T_2-T_1=\frac{2}{k}\left(\frac{x-1}{x+1}\right)+\frac{2}{3k}\left(\frac{x-1}{x+1}\right)^3+\frac{2}{5k}\left(\frac{x-1}{x+1}\right)^5+\dots \tag{8}$$

For temperature differences encountered, x is nearly unity, 1.04 for 50 F difference. The third, fifth, and higher order terms are insignificant and may be ignored. Lastly, letting $2/k$ equal the K of equation 4

$$T_2-T_1=K\left(\frac{x-1}{x+1}\right)=K\left(\frac{R_2-R_1}{R_2+R_1}\right) \tag{9}$$

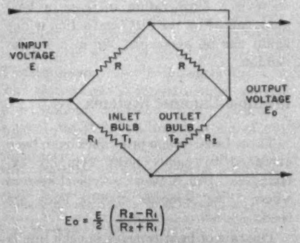

INPUT VOLTAGE E
R
R
OUTPUT VOLTAGE E_0
INLET BULB T_1
OUTLET BULB T_2
R_1
R_2
$E_0=\frac{E}{2}\left(\frac{R_2-R_1}{R_2+R_1}\right)$

Fig. 2. Basic computing bridge

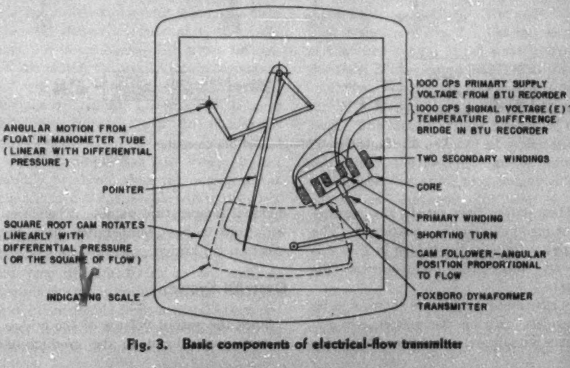

ANGULAR MOTION FROM FLOAT IN MANOMETER TUBE (LINEAR WITH DIFFERENTIAL PRESSURE)
POINTER
SQUARE ROOT CAM ROTATES LINEARLY WITH DIFFERENTIAL PRESSURE (OR THE SQUARE OF FLOW)
INDICATING SCALE
1000 CPS PRIMARY SUPPLY VOLTAGE FROM BTU RECORDER
1000 CPS SIGNAL VOLTAGE (E) TO TEMPERATURE DIFFERENCE BRIDGE IN BTU RECORDER
TWO SECONDARY WINDINGS
CORE
PRIMARY WINDING
SHORTING TURN
CAM FOLLOWER—ANGULAR POSITION PROPORTIONAL TO FLOW
FOXBORO DYNAFORMER TRANSMITTER

Fig. 3. Basic components of electrical-flow transmitter

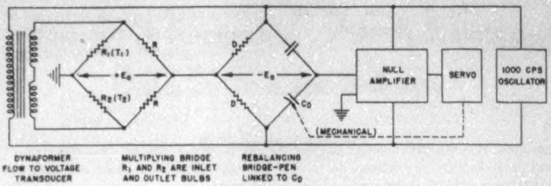

$R_1(T_1)$
R
$+E_0$
$R_2(T_2)$
R
D
$-E_0$
D
C_D
NULL AMPLIFIER
SERVO
1000 CPS OSCILLATOR
(MECHANICAL)
DYNAFORMER FLOW TO VOLTAGE TRANSDUCER
MULTIPLYING BRIDGE R_1 AND R_2 ARE INLET AND OUTLET BULBS
REBALANCING BRIDGE-PEN LINKED TO C_D

Fig. 4. Basic Btu computer

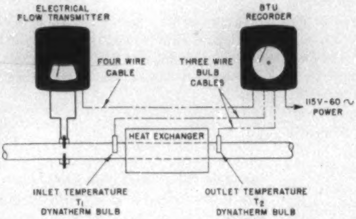

ELECTRICAL FLOW TRANSMITTER
BTU RECORDER
FOUR WIRE CABLE
THREE WIRE BULB CABLES
115V-60 ~ POWER
HEAT EXCHANGER
INLET TEMPERATURE T_1 DYNATHERM BULB
OUTLET TEMPERATURE T_2 DYNATHERM BULB

Fig. 5. Basic Btu computer installation

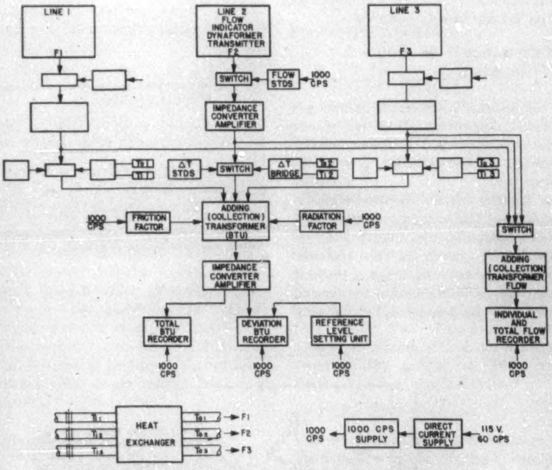

LINE 1
F1
LINE 2 FLOW INDICATOR DYNAFORMER TRANSMITTER F2
LINE 3
F3
SWITCH
FLOW STDS
1000 CPS
IMPEDANCE CONVERTER AMPLIFIER
Ti1
To1
ΔT STDS
SWITCH
ΔT BRIDGE
To2
Ti2
To3
Ti3
1000 CPS
FRICTION FACTOR
ADDING (COLLECTION) TRANSFORMER (BTU)
RADIATION FACTOR
1000 CPS
SWITCH
ADDING (COLLECTION) TRANSFORMER FLOW
IMPEDANCE CONVERTER AMPLIFIER
TOTAL BTU RECORDER
DEVIATION BTU RECORDER
REFERENCE LEVEL SETTING UNIT
INDIVIDUAL AND TOTAL FLOW RECORDER
1000 CPS
1000 CPS
1000 CPS
1000 CPS
Ti1
Ti2
Ti3
HEAT EXCHANGER
To1
To2
To3
F1
F2
F3
1000 CPS
1000 CPS SUPPLY
DIRECT CURRENT SUPPLY
115 V. 60 CPS

Fig. 6. Typical multiple-channel Btu computer

These bulbs then satisfy the requirement made of them in equation 4. The output voltage of the bridge in Fig. 2 is proved linear with temperature difference.

Furthermore, equation 6 shows x to be dependent only on the temperature-difference magnitude, independent of the individual temperature values. Therefore, variations in base temperature have no effect on the bridge.

Over-All System

Since the output voltage of the bridge is now proportional to the product of flow and temperature difference, all that is required to complete the system is an instrument that will measure and record this voltage and supply the 1,000-cps operating voltage to the flow-to-voltage transducer transmitter.

These requirements are met by a standard electronic recorder. Fig. 4 shows, in abbreviated schematic, the over-all system.

The multiplying bridge described in the foregoing is connected so that its output voltage is in series with the output voltage of a second bridge, one arm of which is servo operated.

When these two bridges have equal and opposite output voltages, these voltages subtract, and the input to the null amplifier is zero. This produces no servo action. The position of capacitor C_d, to which the pen of the recorder is connected, represents the rate of Btu transfer. When this rate changes, the amplifier input is no longer zero. The servo system then repositions capacitor C_d to a new point to re-establish a null and to record the new Btu level.

Since the 1,000-cps operating voltage from the oscillator of the instrument feeds all measurement functions, all values are referred to it, and no particular voltage-stabilization requirement is imposed.

A mechanical or electrical integrator operating from the pen position of the recorder provides the continuous solution of equation 2. Fig. 5 shows the basic instrumentation in pictorial form.

On actual heat-exchanger installations several other factors, not previously mentioned, may warrant consideration. If the heat exchanger is so designed that all of the heat removed from one substance is transferred to the other, then the described system suffices. However, if a nonnegligible amount of heat is lost by extraneous conduction or radiation, corrections can be incorporated in the computer either automatically or by means of a manual adjustment.

Similarly, since the cooling fluid suffers a pressure drop through the exchanger, there is a temperature increase, a heat gain, caused by the frictional action which again can be accounted for in a similar fashion.

Multiple-Channel Systems

These basic design principles have been extended to vastly more complex exchangers. Fig. 6 shows a typical system where three separate cooling lines are used.

The rate of heat transfer in each line is converted to a 1,000-cps voltage as previously described. By means of a unique adding transformer,[1] these voltages are accurately summated, and the total Btu rate recorded. The design of this transformer completely eliminates the disturbing interaction among windings found in conventional transformers.

Incorporated in the system is a set of calibration standards. By means of selector switches, 1,000-cps reference-voltage values corresponding to 0%, 50%, 75%, and 100% flow can be injected into the system. Similarly, stable resistance values corresponding to 0%, 50%, 75%, and 100% temperature difference can be inserted to provide means for complete system calibration.

Manual adjustments are provided for the insertion of correction factors for frictional heat gains and radiation losses.

Where variations in heat-transfer rates must be very closely observed, a deviation recorder and a reference-level setting unit are provided. This recorder then shows very small variations from the reference level on a greatly expanded scale.

A flow recorder is included which can be selectively switched to record individual flows or the total of all three. Master-direct and 1,000-cps supply units complete the system. Since the complete system is fed from the 1,000-cps supply, and again referenced to it in the final recorders, the requirement for strict voltage stabilization is eliminated.

Conclusions

As processes move relentlessly to more completely automatic operation, there will be an increasing demand for special-purpose computers. Manual operations have already been greatly reduced by the incorporation of automatic equipment and controls. Inevitably, the processes in turn speed up. Routine human decisions, based upon end results, are required at appalling rates. Performance information must be available to the operators immediately.

Time prevents the evaluation of conditions from a multiplicity of primary measurements. Advanced data systems, automatic computers, will be required to reduce the data to a few basic parameters describing the process status. Decisions can then be based on these, or even automatic control effected from them.

The Btu computer, particularly the multiple-channel unit, is a typical example of such a system, but both the need and the possibilities are unlimited.

References

1. COMBINING MEASURING CIRCUIT, Horace E. Darling. *U. S. Patent no. 2,590,946*, Apr. 1, 1952.

2. COMPUTING CIRCUITS AND DEVICES FOR INDUSTRIAL PROCESS FUNCTIONS, A. J. Hornfeck. *AIEE Transactions*, vol. 71, pt. I, July 1952, pp. 183–93.

Measurement of the Complex Dielectric Constant of Very-High-Dielectric-Constant Materials at Microwave Frequencies

ISIDORE BADY
NONMEMBER AIEE

THE MEASUREMENT of the complex dielectric constant of very-high-dielectric-constant materials at microwave frequencies is considerably more difficult than the measurement of materials with a low dielectric constant. The reason for this can be best illustrated by consideration of a commonly used method of measuring dielectric constant at microwave frequencies which is called the short-circuited line method. The equipment used is shown in Fig. 1.[1,2] The test sample is placed against a fixed short circuit, and the voltage standing wave ratio (VSWR) and location of the voltage minimum are measured. The complex dielectric constant can be calculated from these measurements. This method is satisfactory for materials where the real dielectric constant is relatively low, but poses serious difficulties for materials where the real dielectric constant is very high. In discussing the reasons for the difficulties, low-loss samples and high-loss samples will be considered separately.

If the sample has low losses, it can readily be shown that the impedance at the left face of the sample will be very low except at discrete, very narrow, bands of frequency. Unless the test frequency happens to fall within these bands, it will be very difficult to make an accurate measurement, as the sample will be almost indistinguishable from a metal short circuit; i.e., the voltage minimum will be located at a distance that is very close to an even multiple of a quarter of a guide wavelength from the front face of the sample, and the VSWR will be virtually unaffected by the losses of the sample. It is of course possible to vary the test frequency so as to find a frequency at which the impedance is high. Accurate measurements may be made at this frequency. However, the procedure of hunting for a suitable frequency is time consuming. A more important consideration is the possibility that a frequency that is suitable for accurate measurements, may be far removed from the frequency at which it is desired to make the measurement.

Another method for the measurement of materials with very high dielectric constant and relatively low losses is described by Powles and Jackson.[3] This is a transmission method. The equipment used is shown in Fig. 2. The magnitude of the electric field in the wave guide to the right of the sample will be very low except at frequencies at which the test sample is a multiple of a half wavelength long. The test procedure consists of determining the frequencies at which there are peaks in the magnitude of the electric field in the waveguide to the right of the sample, and measuring the ratio of the magnitude of the electric field in the wave guide to the right of the sample after the sample is inserted to the magnitude before it is inserted, at these frequencies.

Consider that the sample is very lossy so that it appears almost like an infinite transmission line. In the first method discussed, the impedance at the left face

---

Paper 56-260, recommended by the AIEE Dielectrics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted February 23, 1955; made available for printing November 9, 1956.

ISIDORE BADY is with the Signal Corps Engineering Laboratories, Fort Monmouth, N. J.

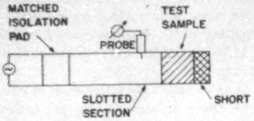


Fig. 1. Equipment used in short-circuit line method

of the sample will be very low at all frequencies. The real part of the dielectric constant can be calculated with fair accuracy from the VSWR. However, it is very difficult to determine the dissipation factor. In the second method, there will be no measurable peaks in the electric field in the wave guide to the right of the sample, and it is, therefore, not possible to make any measurements with this method. A combination of the two methods will permit a measurement of the complex dielectric constant; i.e., measurement of VSWR as per the first method, and the measurement of the ratio of the magnitude of the electric field in the wave guide to the right of the sample after the sample is inserted to the magnitude before the sample is inserted as per second method. This combination is satisfactory provided the test sample is not extremely lossy. In this case the power level in the wave guide to the right of the sample will be so low that it may not be possible to measure it.

This paper describes two methods that have been developed at the Signal Corps Engineering Laboratories which overcome the afore-mentioned difficulties in the measurement of very-high-dielectric-constant materials at microwave frequencies. This paper is intended to be a theoretical presentation of the two methods. A test procedure will be outlined, but there will be no discussion of such practical aspects as losses in the wave-guide walls or other components used. These are discussed in Signal Corps Technical Memorandum *M-1589*,[4] of which this paper is a condensation.

Preparation of samples, including smoothness and parallelism of the front and back faces, and fit of the samples into the wave guide, is extremely important. However, the preparation of the samples will not be discussed here and it will be assumed in the following discussion that the samples are properly prepared.

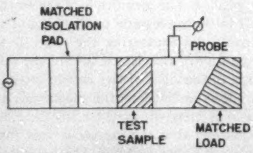


Fig. 2. Equipment used in transmission method

Method A: For Low-Loss Samples

This method is applicable for samples in which the dissipation factor of the material is less than 0.2, and the attenuation of the electromagnetic wave in traveling from one face of the sample to the other is less than 17 decibels. The equipment used is shown in Fig. 3.

The characteristic feature of this method is that the position of the short circuit is varied, and the VSWR in the slotted line is monitored as this is done. If the dissipation factor of the sample is zero, and assuming that the wave-guide and variable short circuit are lossless, the VSWR will be infinite regardless of the position of the short circuit. However, if the sample has a finite dissipation factor, the VSWR will vary, depending on the position of the short circuit, and there will be a minimum and a maximum VSWR. The minimum VSWR will occur when the short circuit is in such a position as to cause the impedance at the left face of the sample to be at its maximum. The reason for this is as follows: The magnitude of the electric field within the sample depends on the impedance that exists at the left face of the sample and is maximum when the impedance is maximum. The amount of energy that is dissipated within the sample is proportional to the square of the magnitude of the electric field within the sample. The greater the energy dissipated within the sample, the smaller the reflected field in the slotted line, and therefore the smaller the VSWR. Similarly, a maximum VSWR will occur at a position of the short circuit at which the impedance at the left face of the sample is minimum.

At any rate, as the position of the short circuit is varied through a half wavelength in the empty guide, the VSWR in the empty wave guide to the left of the sample will go through a maximum and a minimum. It is the purpose of the following analysis to develop the relationship between the maximum and minimum VSWRs and the complex dielectric constant of the sample material.

It should be noted that the approach used in method *A* is similar in some respects to that used in Blackband and Brown in their paper on the measurement of the characteristic impedance and attenuation of cables.[5]

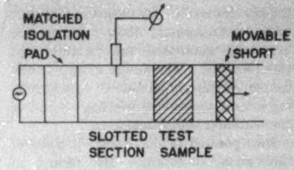


Fig. 3. Equipment used in method A

Derivation of Equations

The first step in the derivation of the equations is to determine the reflection coefficient at the right-hand face of the sample, just inside the sample.

$$\Gamma d(r)=\frac{Zl-Zd}{Zl+Zd}=\frac{jX-Z'd-jFZ'd}{jX+Z'd+jFZ'd} \tag{1}$$

Zl = impedance of the short-circuited section of wave guide terminating the sample; it can be written as jx, since the wave-guide walls are assumed to be lossless

Zd = characteristic impedance of the wave-guide section containing the sample. Since this paper is concerned with very-high-dielectric-constant materials, Zd is also equal (to a very good approximation) to the intrinsic impedance of the sample material. $Zd=Z'd(1+jF)$

$Z'd$ = real part of Zd

F = ratio of the imaginary of Zd to the real part of Zd. It can be shown that $F=\tan\delta/2$ where $\tan\delta$ is equal to the dissipation factor. Since this method is applicable only when $\tan\delta$ is less than 0.2, the maximum value of F that is being considered is 0.1

Through algebraic manipulation, and neglecting terms in F^2, it can be shown that

$$\Gamma d(r)=Ae^{j\psi} \tag{2}$$

where

$$A=\frac{1}{1+F\sin\psi}$$

$$\psi=\tan^{-1}\frac{2Z'dX}{X^2-Z'd^2}$$

It should be noted that A, which is absolute value of $\Gamma d(r)$ is not generally equal to one, as would be the case if the sample material were lossless. If $F=0.1$, A can have any value from 1.11 to 0.91. It appears anomolous that the voltage reflection coefficient in a passive system can be greater than one, as it seems to imply a greater reflected energy than incident energy. However, actually working out the equations will show that the power reflection coefficient is not greater than one and the principle of the conservation of energy is not violated.[illegible]

The second step is to determine $\Gamma d(l)$, the reflection coefficient at the left-hand

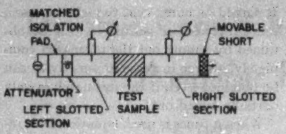
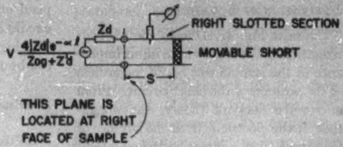


Fig. 4 (left). Equipment used in method B

Fig. 5 (right). Equivalent circuit



face of the sample, just inside the sample.

$$\Gamma d(l)=Ae^{j\psi}e^{-2\gamma l}=Ae^{j\psi}e^{-2\alpha l}e^{-2j\beta l} \qquad (3)$$
$$=AR'e^{j(\psi-2\beta l)}=Re^{j\theta}$$

where

γ = propagation constant in the sample
α = attenuation constant in the sample
β = phase constant in the sample
l = thickness in the sample
$R'=e^{-2\alpha l}$
$R=AR'$
$\theta=\psi-2\beta l$

The third step is to determine Γ_0, the reflection coefficient in the empty wave guide to the left of the sample. The absolute value of Γ_0 is given by

$$|\Gamma_0|^2=\frac{[(D-1)^2-2R\cos\theta(D^2-1)+R^2(D+1)^2]+4FDR\sin\theta}{[(D+1)^2-2R\cos\theta(D^2-1)+R^2(D-1)^2]-4FDR\sin\theta} \qquad (4)$$

$D=Zog/Z'd$, where Zog is the empty wave-guide impedance. $Zog=Zo\lambda og/\lambda o$ where Zo and λo are the free, empty-space impedance and wavelength respectively, and λog is the empty guide wavelength.

In equations 3 and 4 the effect of the losses in the test sample manifests itself in that F is greater than zero and R' is less than one. A careful analysis shows that the error involved in neglecting the terms containing F in equation 4 is less than 0.1%. It can also be shown that for practical cases, the term A shown in equations 2 and 3 can be taken as equal to one with resultant errors in the final calculation of complex dielectric constant of less than 2%. These approximations are discussed in more detail in reference 4.

With the foregoing approximations, the inspection of equation 4 shows that the maximum and minimum values of $|\Gamma_0|$ occur when $\cos\theta=-1$ and $\cos\theta=+1$, respectively.

$$|\Gamma_0|_{\max}=\frac{D-1+R(D+1)}{D+1+R(D-1)} \qquad (5)$$

$$|\Gamma_0|_{\min}=\frac{|D-1-R(D+1)|}{D+1-R(D-1)} \qquad (6)$$

Given the maximum and minimum values of $|\Gamma_0|$, it is straightforward to go to step 4, the determination of $\text{VSWR}_{\max}$ and $\text{VSWR}_{\min}$, the maximum and minimum VSWRs. These quantities are, of course, measured quantities. The final step is to derive the equations which show the relation between the measured quantities, and parameters that are related to the complex dielectric constant of the sample material.

For $\frac{Zo}{Z'd}\frac{\lambda og}{\lambda o}>\coth\alpha l$

$$\frac{Z'd}{Zo}=\frac{\lambda og}{\lambda o}\times\frac{1}{\sqrt{\text{VSWR}_{\max}\text{VSWR}_{\min}}}$$
$$\tanh\alpha l=\sqrt{\frac{\text{VSWR}_{\min}}{\text{VSWR}_{\max}}} \qquad (7)$$

For $\frac{Zo}{Z'd}\frac{\lambda og}{\lambda o}<\coth\alpha l$

$$\frac{Z'd}{Zo}=\frac{\lambda og}{\lambda o}\sqrt{\frac{\text{VSWR}_{\min}}{\text{VSWR}_{\max}}}$$
$$\tanh\alpha l=\frac{1}{\sqrt{\text{VSWR}_{\max}\text{VSWR}_{\min}}} \qquad (8)$$

For $\frac{Zo}{Z'd}\frac{\lambda og}{\lambda o}=\coth\alpha l$

$$\frac{Z'd}{Zo}=\frac{\lambda og}{\lambda o}\tanh\alpha l=\frac{\lambda og/\lambda o}{\sqrt{\text{VSWR}_{\max}}} \qquad (9)$$

The foregoing equations give $Z'd/Zo$ and αl in terms of the measured quantities $\text{VSWR}_{\max}$ and $\text{VSWR}_{\min}$. The complex dielectric constant ϵ^* is derived from $Z'd/Zo$ and αl as follows:

$$\sin\delta=\frac{\alpha\lambda oZ'd}{\pi Zo} \qquad (10)$$

$$\epsilon'=\left(\frac{Zo}{Z'd}\cos\delta\cos\delta/2\right)^2 \qquad (11)$$

$$\epsilon^*=\epsilon'(1-j\tan\delta) \qquad (12)$$

TEST PROCEDURE

The test procedure involves moving the short circuit slowly through a distance of at least a half-guide wavelength, monitoring the VSWR as this is done, spotting the positions of the short circuit where the maximum and minimum VSWRs occur, and measuring the VSWRs at these points. It will be found that the VSWR and the location of the voltage minimum will scarcely change, as the position of the short circuit is changed, except when the short circuit is in a narrow "critical" region. In practice what is done is to set the short circuit at one end of its travel, and set the probe at a voltage minimum. If the short circuit is initially not at a "critical" region, moving it will not appreciably affect the location or magnitude of the voltage minimum. The meter connected to the probe output will thus remain constant, and it is not necessary to monitor the VSWR. However, when the short circuit enters the "critical" region, the probe output will change markedly. The short circuit is then moved in small successive increments, and the maximum and minimum VSWRs determined. The VSWR's involved in this measurement will generally be very large. However, large VSWR's can be accurately measured by the method based on the sharpness of the voltage minimum.[5]

Method B: For Very Lossy Samples

This method is applicable for samples in which the attenuation of the electromagnetic wave traveling from one face of the sample to the other is relatively high; e.g., 17 decibels or greater. The equipment used is shown in Fig. 4.

The basic characteristic of this method is that the movable short circuit is adjusted so that the empty-wave-guide section to the right of the sample becomes resonant, with a resonant rise in the electromagnetic field in the cavity to the right of the sample. This serves greatly to increase the sensitivity of the measurement, and makes possible measurement of very lossy samples that might not otherwise be measurable. In particular, this method has much more sensitivity than the method described previously, in which the wave guide to the right of the sample was terminated in a matched load. If necessary, sensitivity can be further increased by adding a lossless tuner (not shown in Fig. 4) between the left-hand slotted section and the sample.

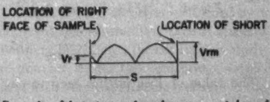


Fig. 6. Variation of voltage to right of sample

DERIVATION OF EQUATIONS

Consider Fig. 4. Due to the high attenuation in the sample, the input impedance of the sample will be Zd, regardless of the position of the short circuit. From microwave network theory, the magnitude of the electric field at the left face of the sample, Vs, is given by the equation:

$$Vs=\frac{2V|Zd|}{|Zog+Zd|}=\frac{2V|Zd|}{|Zog+Z'd+jZ''d|} \qquad (13)$$

where V is the magnitude of the incident electric wave. Since very-high-dielectric-constant materials are being considered, Zd is very much less than Zog, and the quadrature term $jZ''d$ can be neglected.

For convenience, the term "voltage" will be used here in place of "magnitude of the electric field" or "magnitude of the electric wave."

The voltage at the left face of the sample will propagate rightwards, and will, of course, be attenuated. At the right face of the sample, the magnitude of the incident voltage will be

$$\frac{2V|Zd|e^{-\alpha l}}{Zog+Z'd}$$

From microwave network theory, the sample and the entire network to its left can be replaced by a signal generator with a voltage of

$$\frac{4V|Zd|e^{-\alpha l}}{Zog+Z'd}$$

in series with an impedance Zd. This equivalent generator and its impedance is terminated by a short-circuited section of wave guide, as shown in Fig. 5.

The impedance of the short-circuited section of wave guide is $jZog \tan \beta os$ where βo is the propagation constant in the empty wave guide. (The empty wave guide and the short circuit are both assumed lossless.) The voltage at the plane located at the right face of the sample is therefore given by the equation

$$Vr=\frac{4V|Zd|e^{-\alpha l}}{Zog+Z'd}\times\frac{jZog \tan \beta os}{|Zd+jZog \tan \beta os|} \qquad (14)$$

The variation of the voltage amplitude to the right of this plane is as in Fig. 6.

The maximum voltage, Vrm, can be determined by dividing the above expression by $\sin \beta os$.

$$Vrm=\frac{4V|Zd|e^{-\alpha l}}{Zog+Z'd}\times\frac{jZog \tan \beta os}{|Zd+jZog \tan \beta os|}\times\frac{1}{\sin \beta os} \qquad (15)$$

The value of Vrm is thus seen to depend on the distance between the short circuit and the right face of the sample. The highest value of Vrm as s is varied can be shown to be

$$[Vrm]_{\text{highest}}=\frac{4V|Zd|e^{-\alpha l}Zog}{(Zog+Z'd)Z'd} \qquad (16)$$

Adjusting the short circuit to get the highest value of Vrm is equivalent to tuning a series resistance-inductance-capacitance circuit to resonance. Resonance in the wave-guide section to the right of the sample occurs when the reactance in the equivalent network due to the sample is tuned out by the reactance of the wave-guide network to its right, and the only impedance seen by the equivalent generator is $Z'd$.

The maximum voltage in the wave guide to the left of the sample can readily be shown to be

$$Vlm=\frac{2V\,Zog}{Zog+Z'd} \qquad (17)$$

The ratio of the maximum voltage in the left slotted section to the highest maximum voltage in the right slotted section is given by

$$\rho=\frac{2V\,Zog}{Zog+Z'd}\times\frac{(Zog+Z'd)Z'd}{4V|Zd|e^{-\alpha l}Zog} \qquad (18)$$

Therefore

$$\alpha l=\log_e 2\rho\frac{|Zd|}{Z'd} \qquad (19)$$

$Z'd$ can be readily obtained by a measurement of VSWR with the left slotted section. Since the sample has a very high attenuation, the position of the short circuit does not matter in measuring VSWR.

$$\frac{Z'd}{Zo}=\frac{\lambda og}{\lambda o}\,\frac{1}{\text{VSWR}} \qquad (20)$$

In solving equations 19 and 20, it is assumed, as a first approximation, that $|Zd|/Z'd=1$. Values of $Z'd/Zo$ and αl can now be readily calculated. A second (and for all practical cases, final) approximation of the ratio $|Zd|/Z'd$ is now calculated, and more accurate values of $Z'd/Zo$ and αl can now be determined. Equations 11 and 12 are used to calculate the complex dielectric constant from the values of $Z'd/Zo$ and αl.

In the case of extremely lossy material, the electric field may be so greatly attenuated in the sample that it will be difficult or impossible to detect anything in the right-hand slotted section. In this case, considerably more sensitivity can be obtained by placing a tuner between the sample and the left-hand slotted section. It should be noted that though the maximum increase in sensitivity can be obtained by tuning out the mismatch completely, a large increase in sensitivity can be obtained even if the mismatch is only partially tuned out.

When a tuner is used, equation 19 is replaced by equation 21:

$$\alpha l=\log_e(2\rho_t\frac{|Zd|}{Z'd}\sqrt{\frac{Zo}{Z'd}\,\frac{\lambda og}{\lambda o}\times\frac{1}{\text{VSWR}_t}} \qquad (21)$$

VSWR_t is the VSWR obtained when the tuner is used, and ρ_t the corresponding ratio.

TEST PROCEDURE

1. Measure VSWR in left slotted section. Tuner must be removed from equipment in this step.
2. Add tuner if desired. Tune to as low a VSWR as desired, and measure VSWR again.
3. The movable short circuit is set at one end of its travel, the probe in the right slotted section set at a voltage maximum, and the probe output noted. The movable short circuit is then moved in convenient increments and the maximum probe output determined at each step, till the highest maximum output is obtained.
4. The ratio of the maximum voltage in the left slotted section to the highest maximum in the right slotted section is then measured. A convenient way is as follows: First note the highest maximum output in the right slotted section. Take the probe out of the right slotted section and place it in the left slotted section. Increase the variable attenuator till the probe output equals the previous output. The desired ratio can be readily determined from the change in attenuator setting. It is assumed in the foregoing that the response of the probe is the same in both slotted sections, or account is taken of any differences.

References

1. Techniques of Measuring the Permittivity and Permeability of Liquids and Solids in the Frequency Range of 3 c/s to 50 kmc/s, W. B. Westphal. *Technical Report XXXVI*, Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Mass., July 1950.

2. Techniques of Microwave Measurements (book), C. G. Montgomery. McGraw-Hill Book Company, Inc., New York, N. Y. 1947, pp. 625–56.

3. The Measurement of the Dielectric Constant of High-Permittivity Materials at Centimeter Wavelengths, J. G. Powles, W. Jackson. *Journal*, Institution of Electrical Engineers, London, England, vol. 96, pt. III, 1949, pp. 383–89.

4. Measurement of the Complex Dielectric Constants of Very High Dielectric Constant Materials at Microwave Frequencies, I. Bady. *Technical Memorandum M-1569*, Signal Corps Engineering Laboratories, Fort Monmouth, N. J., 1954.

5. The Two Point Method of Measuring Characteristic Impedance and Attenuation of Cables at 3000 mc/s, W. T. Blackband, S. R. Brown. *Journal*, Institution of Electrical Engineers, London, England, vol. 93, pt. III A, 1946, pp. 1383–86.

Dynamic Core Behavior and Magnetic Amplifier Performance

L. A. FINZI
MEMBER AIEE

D. L. CRITCHLOW
ASSOCIATE MEMBER AIEE

ANALYSES of magnetic amplifier systems commonly suffer from a recognized weakness; the magnetic behavior of the cores is approximated by more or less extreme idealizations. Descriptions of amplifier performance therefore, are seldom better than semiqualitative, and those analyses, even though valuable for the understanding of the operation of the various circuits, may not satisfy the more specific needs of the designer who is interested in quantitative predictions. In fact, even without direct experience in the field, one cannot fail to detect a discrepancy between the acceptance of excessively simplified representations of core behavior in the analyses and the need for increasingly refined core-testing procedures that seems to be felt in industrial amplifier production.

Growing concern has been shown in recent times for a detailed appraisal of the relationships between core fluxes and magnetomotive forces which prevail in different manners in different amplifier circuits throughout the dynamic processes of swinging fluxes.[1] It is the purpose of this paper to investigate how some experimental knowledge of core-flux reset, acquired from appropriate magnetic tests,[2] can be used to better describe and predict in detail the operation of certain amplifier systems. Attempts to translate the physical phenomena described into comprehensive analytical formulations are considered premature.

Half-Wave Amplifier Reset by a Constant Magnetomotive Force

The Building Block

Studies of self-saturating multicore amplifiers are initiated frequently by examining the single-core or half-wave system, shown in Fig. 1, as the "building block" of more elaborate systems. The operation during gating half-cycles is described easily with only a moderate knowledge of the magnetic properties of the core. Kirchhoff's voltage equation is written for the gate circuit as

$$v_g = N_g \frac{d\phi}{dt} + (r_g + R_L) i_g + v_\gamma \qquad (1)$$

where v_g is the voltage of the gate power supply, R_L and r_g are the resistances of load and winding, and v_γ is the voltage drop across the rectifier.

The core flux ϕ, commencing from a preset level ϕ_p, rises towards positive saturation at a rate dictated substantially by the applied gate voltage v_g. As saturation is reached, firing occurs and the gate voltage becomes balanced by the voltage drops of gate current. Half-cyclic integration of equation 1 yields

Half-cyclic volt-time area of gate voltage, A_g = swing of gate-winding flux linkages, $N_g(\phi_{max} - \phi_p)$ + half-cyclic volt-time area of gross output (2)

Equation 2 is valid quite generally. It applies for any wave form of the applied gate voltage, e.g., sinusoidal or square waves or in the form of pulses, and also for any kind of magnetic core linked by the winding N_g. In fact, it applies also for any general kind of load, if the lower limit of integration is set at the beginning of flux upswing.

For magnetic amplifiers using core materials with a sharply defined saturation flux level $\pm\phi_s$, it is reasonable to say that the flux linkages of the gate winding cannot increase appreciably beyond the value $N_g\phi_s$, i.e., the swing of flux linkages is expressed as $N_g(\phi_s - \phi_p)$ and equation 2 is rewritten

Half-cyclic volt-time area of gross output = $A_g - N_g(\phi_s - \phi_p) = K_1 + K_2\phi_p$ (3)

where K_1 and K_2, for a given supply voltage, are constants of the amplifier.

In practice, some concern may be shown for small residual inductances that affect the gate winding beyond the postulated saturation of the core. A more precise knowledge of core properties at large magnetomotive forces would be desirable, but this is a matter of refinement in the operation of the half-wave amplifier at ordinary frequencies, and therefore, semiempirical approximations may be adequate for it. Also, some knowledge of the dynamic relation between ascending fluxes and corresponding magnetomotive forces may be of interest, e.g., voltage drops of gate currents, during the prefiring interval of rising fluxes, have some effect upon the rates of rise and therefore upon the time at which firing occurs. But again, approximate criteria with the use of ascending branches of hysteresis loops of sort may be sufficient to evaluate these currents and drops, if one is ever concerned with fine points of this kind. Concluding, quantitative studies of the gating half-cycle do not require any too detailed understanding of the dynamic behavior of the magnetic core.

Such understanding, however, becomes indispensable when evaluating the means to be used to bring, or reset, the core flux from saturation down to the level ϕ_p which establishes the operation of the next gating half-cycle. Mechanical or thermal actions properly applied to the core could be used for this. In all common magnetic amplifiers, of course, electromagnetic means are used. The half-wave amplifier, controlled by virtue of the magnetomotive force of a perfectly constant current in a control winding, has been considered by Uno Lamm as the building block of more elaborate systems.[3] More recent points of view prefer to envisage the reset of core flux as being obtained from the applications of control voltages, which impose negative rates of change to the core flux and swing it down to ϕ_p in the time available between gating half-cycles. In any event, in problems of amplification and control the currents, voltages, and power to be supplied by the signal source over the range of control remain a matter of fundamental importance. This is the delicate point where inaccurate idealizations of the components lead to inaccurate over-all predictions.

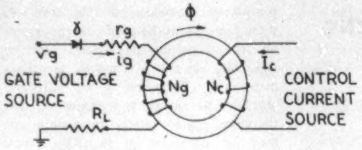
Some Idealizations of the Process of Control by Constant Magnetomotive Force

The half-wave amplifier controlled by a constant magnetomotive force is a useful building block, e.g., for the further study of doubler, center tap, and bridge amplifiers with very high forcing resistors in the control circuit or with signal currents supplied by transistor preamplifiers.

Paper 56-272, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted November 30, 1955; made available for printing October 26, 1956.

L. A. Finzi and D. L. Critchlow are with Carnegie Institute of Technology, Pittsburgh, Pa.

This investigation stems from the work of H. W. Lord[1] and from informal conversations with him on various occasions. Comments by R. W. Roberts and G. F. Pittman, Jr. and also the kind co-operation of Magnetics Inc., in providing most of the cores used are gratefully acknowledged. This work has been performed in part under sponsorship of the U. S. Office of Naval Research.



$$v_g = N_g\frac{d\phi}{dt} + (r_g + R_L)i_g + v_\gamma \qquad F = N_g i_g + N_c I_c$$

Fig. 1. Half-wave amplifier

Present interest in this type of half-wave amplifier is reflected in the activities of a working group on constant current reset within the AIEE Committee on Magnetic Amplifiers. Moreover, methods of core testing and grading are used widely, which to some extent duplicate the operation of this elementary amplifier.[4] In this paper, the mechanism of control by constant signal current receives attention because this is a field in which newly acquired understanding of actual core behavior under certain dynamic conditions can be put to use most basically.

In early times the magnetization curve of the core was often represented by the linearization shown in Fig. 2(A). Accepting this approximation one recognizes that, if the gate current i_g becomes zero before the end of a negative gate-voltage half-cycle, the preset flux ϕ_p, found at the beginning of the next gating half-cycle, is determined by the magnetomotive force $F_c = N_cI_c$, i.e., at the intercept shown in Fig. 2(A) for a case of negative I_c.

Popular misconceptions not withstanding, it is recognized that reversal of the applied gate voltage is not sufficient in itself to stop the flow of positive gate current. One may say here that the same large linearized inductance, which appeared in the control winding throughout the prefiring interval and vanished upon firing, reappears again in the gate circuit as soon as the gate current towards the end of the positive gate-voltage half-wave has decreased to the value i_g'. At this point a "trailing" interval of decaying positive current initiates. The process of reset in this highly inductive circuit now takes place with rates of change substantially equal and opposite to those of the prefiring process of ascending fluxes, and therefore has nearly the same duration. When ϕ has reached the value ϕ_p, no further swing downward is possible, under the idealization, because the rectifier intervenes to block the flow of the negative gate currents that would be needed for this.

Half-wave amplifiers using cores with a square hysteresis loop, however, depart remarkably in their performance from that idealized behavior. For instance, this reasoning indicates that at normal gate voltage, i.e., for half-cyclic gate volt-time area equal to $2N_g\phi_s$, the output at $N_cI_c=0$ is 50% (per cent) of maximum output, while in practice more nearly full output is obtained at zero control.

Various attempts have been made to modify this treatment by using idealized or experimental hysteresis loops of one kind or another to express the instantaneous relationship between $N_gi_g+N_cI_c=F$ and ϕ during reset. It has been recognized, however, that static major loops, e.g., those obtained ballistically or from hysteresigraphs in processes of cyclic magnetization at extremely low frequencies, are not sufficiently meaningful in predictions of transfer characteristics. The effects of eddy currents induced in the laminations have been correctly considered as one reason for the larger values of $|N_cI_c|$ which are generally needed in the experimentation.[5] Additional effects of spin relaxation affect also the dynamic processes of swinging fluxes. These effects are immediately recognizable at low frequencies.

"Dynamic" major loops, e.g., those obtained by applying normal sinusoidal voltage of rated frequency directly at the gate-winding terminals, have been considered; Fig. 2(B). The mechanism of trailing gate currents recognized in the idealized situation of Fig. 2(A) is found again. As the gate current recedes from its peak value, negative rates of change of core flux become appreciable before the end of the positive half-cycle, with i_g departing from the value v_g/r_gR_L. Trailing positive gate currents keep flowing into the negative half-cycle; and during the first part of the negative half-cycle the reset takes place as in equation 1, with $v_\gamma=0$. Therefore, the back of the major loop of sinusoidal fluxes may be accepted tentatively as an approximation of the corresponding part of the minor loop described in the amplifier operation.

At some time in the negative half-cycle i_g becomes zero and the rectifier γ blocks its reversal. However, the flux level ϕ_m thus reached is not necessarily the final ϕ_p; the use of a dynamic loop in a search of the level at which the flux shall presumably remain constant is self-contradictory. Rather, one may predict that, after extinction of the gate currents, the core flux will wander further from ϕ_m towards the level ϕ_n on the static loop and may or may not reach the latter value, depending on the time that remains available for this.

Thus, the level ϕ_p actually attained at the end of the negative half-cycle is indeterminate between the two limits ϕ_m and ϕ_n. The "control curve" relating ϕ_p to N_cI_c lies somewhere within the band between the descending branches of those two major loops. According to equation 3, this control curve is geometrically affine to the transfer characteristic which relates half-cyclic gross output to N_cI_c. The control curve, however, is not an "intermediate loop" if this word is to be preserved to express instantaneous relationships between ϕ and F. In the same way, it may be misleading to use the word permeability for any slope of that curve.

In some situations of amplifier circuitry, i.e., self-saturating amplifiers with low control circuit impedance, it has been shown that knowledge of both the static

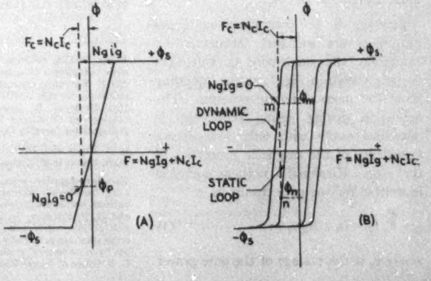


Fig. 2. A—Single-valued linearization. B—Attempt to describe the behavior of amplifier by means of major hysteresis loops

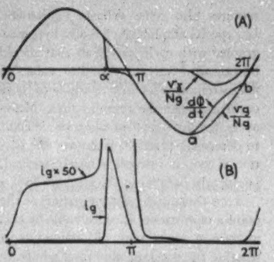


Fig. 3. One cycle of steady-state operation, typical core amplifier, normal voltage = 57 volts rms, 60-cycle, r_g = 15 ohms, R_L = 150 ohms, N_g = 1,100 turns

A—Gate voltage per turn v_g/N_g, reset voltage $d\phi/dt$, and rectifier voltage per turn, v_R/N_g
B—Gate current i_g, two different scales

loop and the dynamic loop of sinusoidal fluxes leads to fair predictions of the control requirements over the whole range of outputs.[6] Here this information is much less conclusive.

EXPERIMENTAL OBSERVATION OF A HALF-CYCLE OF RESET

Fig. 3 shows the processes under investigation in an amplifier of the type discussed at a given value of constant I_c. The ripple observed in I_c during reset is less than 0.5%. The core material is a good grade of commercial grain-oriented 50% nickel-iron alloy. The thickness of laminations is 2 mils, inside diameter is 1.25 inches, outside diameter is 1.75 inches, core mean length is 12 centimeters, and tape height is one inch. A good germanium rectifier with negligible leakage has been chosen to avoid confusion of issues that often arise in such studies. In the experimentation exhibited later, this "typical" core was used for consistency (except for Figs. 10, 11, and 24).

Small negative values of $d\phi/dt$ are noticeable already in the positive half-cycle as i_g recedes from its maximum value. After reversal of the gate voltage at π, small positive gate currents keep trailing through the conducting rectifier, and, as their ohmic voltage drops are soon inappreciable, the rate of change of resetting flux linkages $N_g d\phi/dt$ almost coincides with the gate voltage. At *a* the trailing current dies out and a new process initiates in which the only magnetomotive force acting on the core is $N_c I_c$. During this new process $|N_g d\phi/dt|$ is less than $|v_g|$, and their difference appears as a reverse voltage across the blocking rectifier. Finally, at *b*, the rectifier restrikes into conduction, positive gate currents flow, and $N_g d\phi/dt$ once more becomes nearly equal to v_g.

Tentatively, the words "backfiring" and "robbing" have been used in the art to designate the processes from π to *a* or from *b* to the end of the negative half-cycle. One could also say that these are intervals in which, as the rectifier is conducting, the reset is "produced" by the gate voltage. But really, as seen from equation 1, the rate of change of flux linkages is larger in absolute value than the gate voltage, even if only by the generally small amount of ohmic voltage drops of positive gate current. Also, the power involved in the reset is supplied from the control circuit during these processes, even though the total energy per cycle $\int_0^T I_c N_g d\phi/dt + dt$ is zero. Actually, the gate circuit is acting as a biased-rectifier limiter, with the time-varying bias v_g. It appears more appropriate to name these intervals "gate-limited," as distinguished from the "constant-mmf (magnetomotive force)" interval *ab* of zero gate currents.

Experimental evidence of this sort is familiar. For different cores, gate voltages, and circuit parameters the oscillograms may vary, with varying relative significance of the two different processes of gate-limited and constant-mmf reset. Also, in some cases, restriking of the rectifier occurs more than once in the half-cycle and, accordingly, additional intervals of gate-limited reset appear between intervals of constant-mmf reset. (A situation of this type is shown later in Fig. 10.) In any case, the total cyclic gross output, i.e., the cyclic volt-time area of voltage drops of large or small gate currents through load, gate-winding resistance, and forward conducting rectifier throughout the cycle in the steady state, is equal to the volt-time area of the blocking rectifier. This is an immediate result of a time integration of the general equation 1 over one cycle.

FLUX REVERSAL BY CONSTANT MAGNETOMOTIVE FORCE

The rate of change of flux of a core receding from positive saturation under the action of a constant negative magnetomotive force F_K has been studied experimentally by Huhta.[2] It has been described further by Friedlaender[7] with a mathematical model in which the motion of domain walls enclosing the growing regions of reverse magnetization is expressed (for grain-oriented materials) by a nonlinear differential equation of the first order. This accounts for the effects of eddy currents and spin relaxation. Also, test methods are described by Roberts,[4] as used for industrial purposes of core grading and matching, in which the core is subjected jointly to the magnetomotive force of a half-wave sinusoidal current source which drives the core periodically far into positive saturation, and to the negative magnetomotive force F_K of a constant current. F_K is greatly overruled by the saturating magnetomotive force essentially over the whole positive half-cycle, but it provides the incentive to reset from positive saturation during the remaining time in the cycle. The method is directed primarily to over-all evaluations of flux swings read out on an averaging voltmeter, but oscilloscopic recordings of instantaneous reset voltages which are induced in a pickup coil are easily obtained.

Curves of reset voltage $d\phi/dt$ versus time, obtained by use of the repetitive procedure of reference 4, at 60-cycle repetition rate are shown in Fig. 4 for the

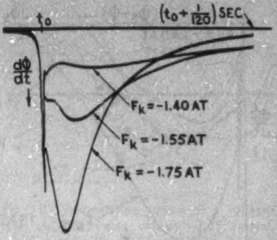


Fig. 4. Oscillograms of core reset voltage versus time, for various values of the resetting mmf F_K

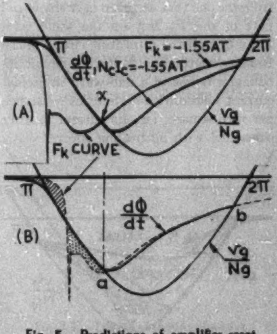


Fig. 5. Predictions of amplifier reset

A—By matching time origins
B—By matching fluxes

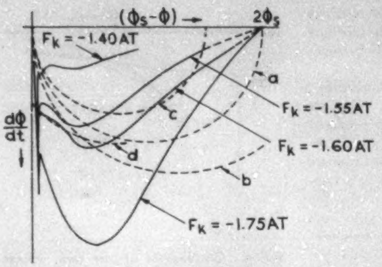


Fig. 6. Voltage versus flux swing curves from F_K tests and elliptical loci of sinusoidal gate supply

a—normal gate voltage, 57-volts rms, 60 cycle
b—1.25 normal
c—0.75 normal
d—normal locus translated for use in Fig. 9

core of the amplifier considered in Fig. 3. These curves show only minor departures from those obtained by the "single-shot" methods of reference 7 with the use of comparable values of saturating magnetomotive force. Single-shot methods, of course, yield information on the complete reset process up to the time at which $d\phi/dt$ finally becomes zero, while the process is not carried beyond the end of the negative half-cycles in the cyclic procedure. In these tests, the peak of saturating field intensity is $H_{max} = +11.4$ oersted, and the slope of the front of the "step" is $dH/dt = -4.3$ oersted per millisecond. The time at which dH/dt becomes zero is indicated by t_0.

Since the amplifier of Fig. 3 during the time interval *ab* of zero gate current is resetting solely by the action of the constant mmf $F_c = N_cI_c$, it appears that the information conveyed by the corresponding F_K test may supply some valuable information on the amplifier behavior during that interval. It must be recognized, however, that the amplifier core during the gate-limited process has changed its magnetic status from the initial positive saturation through a mechanism different from the one that occurred over any initial part of the F_K test on the same core. Even though the saturating circuit of the Roberts' method is a half-wave circuit, its behavior is one of a half-wave sinusoidal current source with high series resistance per turn square, i.e., its current is truly sinusoidal with no trailing effects.

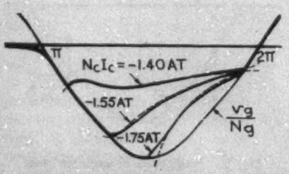


Fig. 7. Predictions of amplifier reset under use of Fig. 6, normal gate voltage

Graphical procedures in which the F_K-test curve for $F_K = F_c$ is superposed on the half-wave of negative gate voltage (per turn), matching the two time origins as shown in Fig. 5(A) and obtaining the time of initiation of the constant-mmf reset at the intercept x of those two curves thus drawn (appealing as they may be in their simplicity) would be misconceived. The manner in which flux reversals proceed in time are totally different and unrelated in the two situations of the F_K test and of the gate-limited amplifier core. The difference between the ordinates of those two curves before they intersect has no physical meaning and therefore its convergence to zero has no meaning either. In fact, at the particular instant determined by that intercept, the amplifier core has reached a flux level which is certainly different from the flux level it reached in the same interval of time in the F_K test. If the amplifier core at the point really becomes submitted solely to the mmf $F_c = N_cI_c$, e.g., by a manual interruption of the gate circuit, its further reset will not even follow the F_K curve thus drawn because, for instance, different fluxes, i.e., different penetrations of the magnetization reversal into the material, imply different conductances for the paths available to the eddy currents.

A more pertinent procedure would be to determine the point of crossover a, i.e., the time at which the trailing gate current extinguishes, from graphical operations performed on the dynamic major hysteresis loop of sinusoidal fluxes. The intercept m of Fig. 2(B) yields approximately the flux ϕ_m reached at the time of zero gate current; thus, the duration of the gate-limited process results from a simple time integration of the gate voltage. But, the experimental determination of such loops requires that the core be provided with a comparatively large number of turns. The most immediate advantage of the F_K-test techniques is that cores can be tested without any preliminary use of winding machines. It requires also some delicate equipment, i.e., special amplifiers[8] or costly hysteresigraphs[9] with cyclic processes that should be extended into saturation at values of magnetomotive force comparable to those occurring when the amplifier fires. Moreover, the intercept m may be difficult to determine precisely whenever the portion of loop of interest is nearly vertical, i.e., nearly parallel to the straight line.

Even though the determination of the point a of crossover is conceivable by this method, a doubt may still remain as to whether the constant-mmf reset will truly proceed from there along any portion of the F_K-test curve. It is recognized that the instantaneous behavior of a core, i.e., the relation between applied magnetomotive force and corresponding rate of change of core flux, depends on the electromagnetic "history" of the core previous to the instant considered. As inertia effects of moving walls appear negligible in grain-oriented nickel-iron alloys for the applications considered here, this history is expressed by geometry (shape and position) of the domain walls enclosing regions of reverse magnetization at that instant. This is not a quantitatively useful concept, since there is no adequate way in general to establish the influence of previous electromagnetic manipulations upon the resultant geometry, nor to decide to what extent differences of geometry are significant in terms of subsequent behavior. To be safe, one should conclude that results of dynamic tests of any kind can be applied only to situations entirely identical to those that prevailed before and throughout those tests; this would be a most frustrating conclusion that may be circumvented only if the

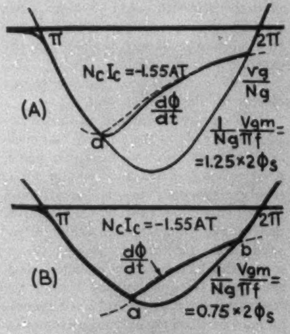


Fig. 8. Amplifier reset

A—Gate voltage 1.25 normal
B—Gate voltage 0.75 normal

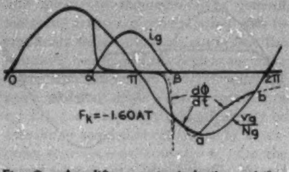


Fig. 9. Amplifier reset, inductive-resistive load, normal voltage

unclear concept of history can be substituted, for the purposes at hand, by some other concept more precisely defined and related to operation of the amplifier.

The quantity most readily available for this is the core flux. In the following, the assumption is made that, for a given core at a certain flux level ϕ at a certain instant, the existence of a certain mmf F determines uniquely the corresponding $d\phi/dt$ or vice versa. This is a bold assumption because identity of wall geometries implies identity of core fluxes in the given core, but the converse is not necessarily true; a common value of core flux may result from a variety of wall geometries. The curves of Fig. 7 of reference 2, seemingly obtained by applying a fixed negative F_K to a core initially brought to some flux level $\phi_0<\phi_s$ by some particular process that ascended partway from initial negative saturation,[10] understandably invalidate this tentative assumption for that case. Nevertheless, the assumption may remain reasonable if, in different processes, the growth of domains from individual nuclei has proceeded to the extent of a lateral merging of single domain walls into a nearly planar wall, or also, approximately, if different processes leading to the same ϕ have not been too dissimilar; i.e., if the reversals are initiated originally from comparable magnitudes of the peak of saturating field intensity, and proceed from there monotonically, with magnetomotive forces varying at moderate rates. The purpose of the experimentation that follows is to determine the practical limits of the validity of the assumption which is thus restricted.

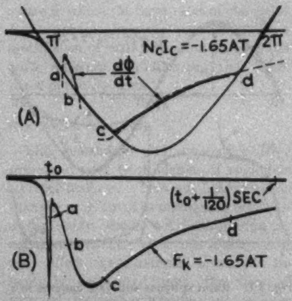


Fig. 10. Reset of amplifier with multiple-robbing core

PREDICTIONS OF AMPLIFIER RESET FROM F_K TESTS

The assumption of uniqueness of the relation between ϕ, F, and $d\phi/dt$ applied to the crossover point a of Fig. 3 implies that a matching of two of these quantities at the boundary between gate-controlled and constant-mmf reset results in the matching of the remaining variable. The procedure mentioned in the previous section, i.e., searching for the point m on the dynamic loop of Fig. 2(B), is substantially one in which values of F and $d\phi/dt$ are matched.

In the following, a procedure is preferred in which a is determined by the matching of values of fluxes and flux rates; the information supplied by the F_K tests is completely sufficient for this. This matching of boundaries results in the "equal-areas criterion" used in Fig. 5(B); the crossover a is obtained by superposing the voltage-versus-time curve of the F_K test with $F_K=N_cI_c$ on the v_g/N_g curve and by translating it horizontally until the dashed and the crosshatched areas become equal.

A comparison of the predictions of Fig. 5(A) with oscillograms of actual amplifier reset indicate that the arbitrary matching of time origins yields erroneous results in this case by -20%. The procedure of Fig. 5(B) is in more reasonable agreement with the actual amplifier behavior. Evidently, the "robbing" effects of the initial gate-limited processes cannot be judged directly on the basis of the shape of the volt-time curve of the F_K test in its early parts; curves with initial high values of $d\phi/dt$ followed by a quick decay do not necessarily lead to lesser half-cyclic flux swings than curves which, for the same value of F_K, may appear more nearly sinusoidal. In the example of Fig. 5, gate-limited intervals are quite extended through the reset half-cycle, yet the volt-time area of reset is very close to the half-cyclic volt-time area that would be read out from the averaging rectified voltmeter of Roberts' industrial test. The feared mutilation by robbing really turned out to be rather harmless. For more extensive studies of the control ampere-turns needed by a given amplifier for various gate voltages, wave forms, and frequencies, and also for various types of loads a more systematic procedure can be used. For this, the voltage-versus-time curves of the F_K tests (shown in Fig. 4 for the core of this amplifier) are converted into the voltage-versus-flux swing curves shown in Fig. 6. If accurate integrating circuits are available, the new curves can be recorded directly during the F_K test. The curves shown in Fig. 6 result from graphical integrations of the curves of Fig. 4. On the other hand, the locus of gate voltage versus flux swing during the gate-limited process is also plotted in Fig. 6. If one neglects the small flux swing that may have taken place already before initiation of the negative half-cycle, this locus is $d\phi/dt=$ function of $1/N_g\int_{\pi/\omega}^{t} v_g\,dt$. For sinusoidal gate voltages with crest V_{gm}, the locus is an arc of ellipse with ver-

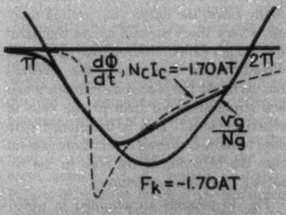


Fig. 11. Reset of amplifier with round-knee grain-oriented core material

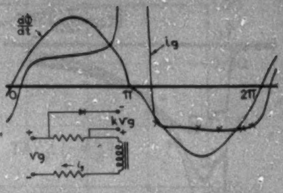


Fig. 12. Magnetizing current of a voltage-reset amplifier, gate voltage = 0.75 normal, k = 0.1. Predictions from Fig. 6

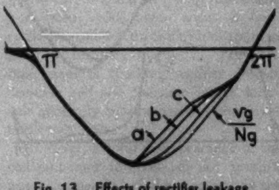


Fig. 13. Effects of rectifier leakage

a—R_{shunting} = ∞ ohms
b— " = 250 kilohms
c— " = 62 kilohms

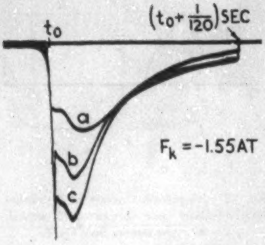


Fig. 14. F_K curves for different values of initial saturating H_{max}

a—saturating $F=90.00-1.55$ AT, $H_{max}\cong$ 10.00 oersted

b—saturating $F=6.00-1.55$ AT, $H_{max}\cong$ 0.47 oersted

c—saturating $F=4.00-1.55$ AT, $H_{max}\cong 0.26$ oersted

tical axis V_{gm}/N_g. The horizontal axis initiates at the origin of the abscissae and terminates at $\phi_s-\phi=V_{gm}/\pi\int N_g$. Proper choice of scales converts these ellipses into circles for a fixed value of the frequency. The first intercept of this locus with the pertinent core curve yields the value of gate voltage per turn at which crossover occurs and therefore also the initial point of the section of the pertinent F_K curve of Fig. 4, which shall approximate the $d\phi/dt$ values of the amplifier core during its further process of constant-mmf reset.

Fig. 7 compares experimental oscillograms of half-cyclic reset for the amplifier, described previously, at normal 60-cycle sinusoidal gate voltage, with predictions obtained with the use of F_K-tests information by the procedure of Fig. 6. The ratios between volt-time areas of reset $N_g(\phi_s-\phi_p)$ and half-cyclic volt-time area of gate voltage A_g are 0.452, 0.712, 0.883 for N_cI_c equal to -1.40, -1.55, and -1.75 respectively. (This information is easily converted into points of a transfer characteristic with the use of the basic equations 2 or 3, given initially; as summed up later in Fig. 23.)

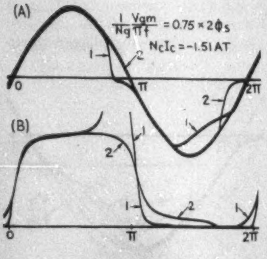


Fig. 15. Reset voltages and gate currents in a jump-down phenomenon

1—before triggering

2—after triggering

Fig. 8 extends this comparison to predictions of half-cyclic reset with gate voltages larger and smaller than the normal value. Checks performed at 400-cycle gate frequency yield the same kind of agreement if the amplifier windings are progressively wound to minimize capacitive effects. The same procedure applies also to other forms of gate voltage, (the gate locus of Fig. 6, of course, then ceases to be an ellipse) and also to situations of nonresistive loads. Fig. 9 refers to a case of resistive-inductive loading without a free-wheeling rectifier. The gate current upon firing is dictated substantially by load and gate voltage and flows with comparatively large values well into the negative half-cycle. Small positive trailing currents keep flowing past the apparent angle of extinction β, i.e., the angle at which the gate current has decayed practically to zero in terms of load output, in the same way as trailing currents extended past $\omega t=\pi$ in the previous cases of resistive loads. The gate-limited process is different only because the gate-limiting voltage is expressed by a different portion of its sine wave. The equal-areas criterion still applies, and the procedure of Fig. 6 can be used with a shift of the elliptical locus (gate voltage per turn versus flux swing) to the left, as seen in curve *d* of Fig. 6, to account for the fact that the gate-limited process of resetting core flux initiates practically at $v_g=V_{gm}$ sin β. For any given load, the angle β can be predicted from nomograms based on approximate methods of switched-circuit analysis in terms of the angle of firing α, taken as the independent variable.[11] Unfortunately, in this problem α is not immediately known in terms of N_cI_c and therefore predictions of performance imply tentative cut-and-try procedures. In many cases, the initial gate-limited process is very short because of the comparatively high gate-limiter voltage v_g existing already at the beginning of the reversal, but certainly the procedure remains much less straightforward than in the case of resistive loads. This is mentioned here only to show that even in this more general case the behavior of the core in the resetting process is predictable.

Fig. 10 presents an amplifier in which the half-cyclic reset occurs with three gate-limited intervals separated from one another by two constant-mmf intervals. The prediction of the end of the first gate-limited interval *a* is obtained as before; the subsequent constant-mmf interval, described by a properly matched portion of the F_K curve, terminates at the point *b* at which $d\phi/dt=v_g/N_g$. Here the rectifier γ restrikes into conduction and the reset becomes gate-limited. However, at *c*, a new constant-mmf interval initiates; the new crossover *c* is found by the equal-areas criterion. The unknown point *c* results from two simultaneous conditions: 1. identity of amplifier gate voltage per turn and of instantaneous $d\phi/dt$ on the pertinent F_K curve, and 2. identity of volt-time area of amplifier reset per turn and of F_K-curve volt-time area. The procedure of Fig. 6 can be used again but now the gate locus must be translated to the left by an amount which takes into account the fact that the amplifier in the interval *ab* has been reset somewhat less than it would have been in the case of an uninterrupted gate-limited process. In other words, to determine *c* the gate locus must be translated to the left by an amount equal to the volt-time area per turn of the blocking rectifier in the intermediate constant-mmf interval *ab*.

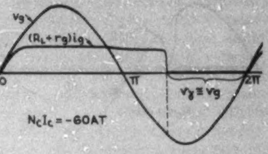


Fig. 16. Operation far in the negative branch of the characteristic

Further application of the equal-areas criterion is shown in Fig. 11 for an amplifier with "round knee" 50% nickel-

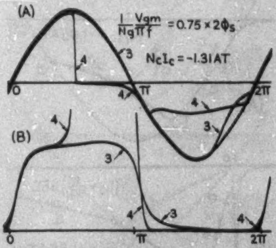


Fig. 17. Reset voltages and gate currents in a jump-up phenomenon

3—before triggering

4—after triggering

iron alloy. In themselves, the general matching criteria suggested apply also to amplifiers with "48-alloy" cores, but the matching here should not neglect the more appreciable volt-time area of amplifier reset gathered already before reversal of v_g; i.e., the gate locus of Fig. 6 should be translated somewhat to the right to express that $(\phi_s-\phi)>0$ at $v_g=0$. On the other hand, the criteria do not appear satisfactory for amplifiers with Mu Metal cores; the mechanism of flux reversal here seems to be significantly different, so as to void the basic assumption which was introduced at the end of the section on flux reversal.

The validity of the assumption, for grain-oriented 50% nickel-iron alloys, can be recognized also in situations quite different from those related to constant-current control. Fig. 12 presents a comparison of calculated and measured instantaneous magnetizing current for a process of asymmetric cyclic magnetization, substantially identical to the reset of a Ramey amplifier controlled by a half-wave rectified sinusoidal voltage. The calculated crossed points were obtained from the ellipitical locus of Fig. 6 for 75% normal gate voltage by stating that if $d\phi/dt$ and $\phi_s-\phi$ are known at any instant (as dictated by the applied voltage and thus related by the ellipse) the corresponding value of $F=N_g i_g$ is the one pertaining to the voltage-flux swing curve of the core which passes through that point of the ellipse.

This reveals that the F_K-test curves express most significant information on the dynamic core properties, beyond the specific problem of constant-current reset; for instance, descending branches of dynamic loops can be calculated from the results of the F_K tests. However, attempts to determine instantaneous values of ϕ for sinusoidally varying F yield poorer accuracy. Again, this seems to be a case in which the large rates of change of F throughout the cyclic process impose a mechanism of wall formation and growth quite different from the one that takes place in the F_K tests.

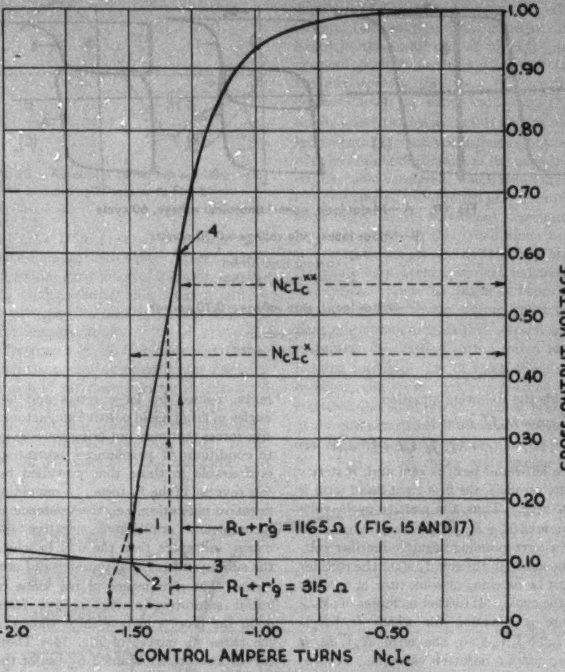


Fig. 18. Bistable behavior at low outputs

Rectifier Leakage

The reasoning and experimentation presented refer to situations of negligible rectifier leakage. This has been done so that the problem of dynamic core behavior under study would not be obscured by extraneous effects. Rectifier leakage does not affect gate-limited processes inherently associated with the flow of forward rectifier currents, but leakage currents i_{ge}, during intervals of reversed rectifier voltage, modify the concept of constant-mmf reset. The magnetomotive force applied to the core after crossover is no longer $F_c=N_cI_c$ but rather $F=N_cI_c+N_g i_{ge}$, and thus $d\phi/dt$ at any instant is expressed by a point on that F_K curve for which $F_K=F$. On the other hand, the leakage current i_{ge} depends on the difference between gate voltage v_g and reset voltage $N_g d\phi/dt$. Therefore, in the absence of simple analytical expressions for both the F_K-test curves and the volt-ampere characteristic of the rectifier, step-by-step procedures (applied over comparatively short time intervals Δt, etc.) are needed to describe this reset process. As the mechanism is clear, the choice of specific procedures or the introduction of simplified empirical criteria becomes a matter of individual experience; a practical limitation may be found in the need for interpolations between the discrete number of F_K curves available. Effects of rectifier leakage are simulated qualitatively in Fig. 13 by use of resistances, shunting a germanium rectifier. Rectifier leakage aids the reset of the amplifier considered at any given N_cI_c; however, this aid is most pronounced at comparatively small absolute values of N_cI_c, (large rectifier volt-time areas). Thus, rectifier leakage has an unfavorable effect upon the steepness of the transfer characteristic of the half-wave amplifier, as generally known.

Region of Minimum Output

Absolute values of $|N_cI_c|$ increasing from zero yield larger areas of half-cyclic reset and thus diminish outputs as seen, for instance, in Fig. 7. The volt-time areas of blocking rectifier voltage decrease and also the gate-limited intervals generally become more predominant in the half-cycle.

However, a total disappearance of the constant-mmf intervals is not possible at any value of the control current in the steady state, not even during the condition of "under excitation." Integration of equation 1 for the duration of one cycle

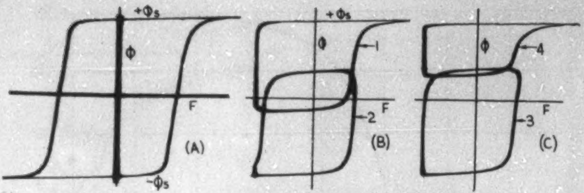


Fig. 19. A—Major loop, normal sinusoidal voltage, 60 cycle

B—Minor loops, gate voltage = 0.70 normal

1—before jump down
2—after (see Fig. 15)

C—Minor loops, gate voltage = 0.70 normal

3—before jump up
4—after (see Fig. 17)

yields the following equation

$$\int_0^T v_g dt = N_g \int_0^T d\phi/dt\, dt + (r_g+R_L)\int_0^T i_g dt + \int_0^T v_\gamma dt \quad (4)$$

The left-hand term is zero and, if steady state prevails, the first right-hand term is also zero. Thus, the positive-cyclic volt-time area $(r_g+R_L)\int_0^T i_g dt$ must be balanced by a corresponding negative-rectifier volt-time area; in the steady state the rectifier must be blocking at some time or another in the cycle. If further increases of $|N_cI_c|$ cause a situation for which $(r_g+R_L)\int_0^T i_g dt + \int_0^T v_\gamma dt > 0$, the term $N_g\int_0^T d\phi/dt\, dt$ acquires a negative value, i.e., a downward net swing of the core flux takes place in any such cycle.

The F_K-test curves used so far do not supply adequate information in the region of low outputs of the transfer characteristic, because small crests of output currents, caused by large resets and late angles of firing, yield positive magnetomotive forces that may not bring the core into conditions of pronounced saturation, comparable to those that prevailed before reversal in the F_K tests. A remaining negative nucleation, i.e., the existence of incompletely annihilated negative domains, influences then the way in which the subsequent reversal initiates and procedes. The significance of the value of initial saturating magnetomotive force upon the process of reversal is shown in the three F_K curves of Fig. 14. These curves have been obtained by use of the cyclic procedure of the Roberts' method, all for a common value of the step of resetting mmf F_K. In curve *a* the peak of the sinusoidal saturating half-wave corresponds to a resultant field intensity $H_{max} \cong +10.0$ oersted. Curves *b* and *c* have been obtained by truncating the top of that half-wave with a current limiter, so that H_{max} has been reduced to +0.47 and +0.26 oersted respectively, without changing the front of the resetting step F_K. No significant departures from curve *a* result until $H_{max} < 2.0$ oersted.

The well-known jump or triggering phenomena which are observed at two different values of N_cI_c in the region of low output of the transfer characteristic can now be explained. Consider the situation marked *1* in the oscillograms (A) and (B) of Fig. 15. This is a steady-state condition reached by gradual increases of control up to a certain value $|N_cI_c^*|$ for which firing occurs rather late in the positive half-cycle. The gate voltage is less than normal, its half-cyclic volt-time area being equal to $0.75 \times 2N_g\phi_s$. With a total circuit resistance, corrected to account for forward rectifier drops, of 1,165 ohms, the peak of output current upon firing provides a saturating field intensity of +2.5 oersted.

If in these conditions an incremental increase is given to $|N_cI_c^*|$, e.g., at the beginning of a positive half-cycle, the prefiring gate current in the upward flux swing is somewhat larger; its ohmic drops are also larger and firing occurs at a somewhat later time in the positive half-cycle, with accordingly lower values of peak output current. The flux reversal now takes place at a faster rate than in the previous cycle. In fact, during gate-limited intervals the voltages $(r_g+R_L)i_g$ are larger for two reasons: 1. the increase of $|N_cI_c|$ in itself calls for larger values of the trailing currents, and moreover 2. the requirements of core magnetomotive force are reduced by the initial presence of

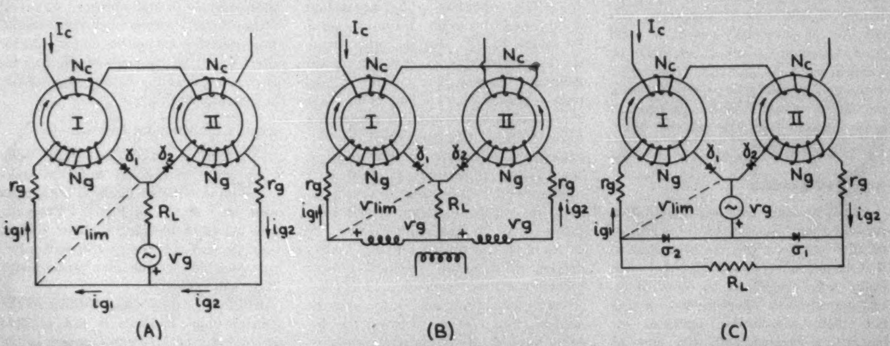


Fig. 20. Two-core amplifiers

A—Doubler B—Center tap C—Bridge circuit

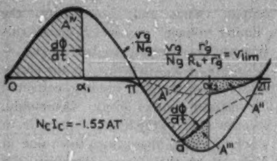


Fig. 21. Reset of a doubler amplifier. Matched typical cores; normal gate voltage, $R_L=300$ ohms, $r_g'\cong125$ ohms

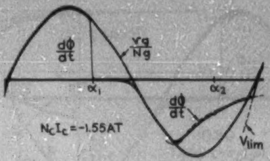


Fig. 22. Reset of a center tap amplifier, same building blocks as in Fig. 21

more appreciable residual negative nucleation, i.e., the descending branch of the minor loop is closer to the $F=0$ axis. In some cases, an interrupted gate-limited process may cover the whole reset half-cycle, and the rectifier never blocks. In other cases, constant-mmf intervals are still observed, but the increase of $|N_cI_c|$ and also the general change of the shape of the pertinent F_K curve result in a pronounced reduction of the rectifier volt-time area. In all events, the flux level reached at the end of the cycle is lower than the initial ϕ_p. The subsequent upswing terminates in an even more delayed firing with larger residual nucleation, etc., and it is likely that the transient process of a ratcheting down of the minor loop may continue for quite a few cycles. A steady state may be reached conceivably with a minor loop "floating" midway, without reaching either $+\phi_s$ at its top or $-\phi_s$ at its bottom. For reasonably square materials with inside-diameter and outside-diameter ratios reasonably close to unity the ratcheting process terminates at a minor loop which at its bottom is very close to $-\phi_s$. The traces marked with *2* in Fig. 15(A) and (B) describe the new steady state. The transfer characteristic of Fig. 18, obtained by increasing values of $|N_cI_c|$, shows a jump of output at $|F_c|=|N_cI_c^*|$. This jump is not very conspicuous but is still noticeable.

Further increases of $|N_cI_c|$ make i_g larger both in the upswing and downswing. The cyclic output increases, although no firing occurs, and the bottom of the loop becomes anchored to the level $-\phi_s$. If $|N_cI_c|$ is increased further, the familiar operation over the negative branch of the characteristic takes place as shown in Fig. 16. Here the core flux is $-\phi_s$ at the beginning of the positive half-cycle; the gate current rises from zero, as given by $i_g\cong v_g/(r_g+R_L)$, until its magnetomotive force is large enough to make flux rises possible. When the positive i_g decreasing from its crest is no longer sufficient to sustain i_g, a process of trailing currents initiates, and the flux resets from its top value to $-\phi_s$. As negative saturation is approached, the negative gate voltage or a large part of it appears as a reverse voltage across the rectifier. Thus, a constant-mmf interval initiates, but the flux change during this interval is very small since the core is already close to negative saturation.

Decreases of $|N_cI_c|$ toward zero move back the operating point on the negative branch of the characteristic at least until the control again reaches the value $N_cI_c^*$. The point thus reached is not necessarily one of unstable operation now. Incremental decreases of $|N_cI_c|$ and associated decreases of i_g cause a decrease of cyclic output, but no significant changes are found in the conditions of already pronounced positive nucleation from which the flux upswing initiates. Further decreases, down to some value $|N_cI_c^{**}|$, finally lead to an incompatible situation of incremental decreases of output associated with incremental increases of the negative-rectifier volt-time area, which result from less pronounced gate-limited processes and accordingly more significant constant-mmf intervals. In this conflicting situation, a transient process initiates in which the minor loop ratchets upward in successive cycles. Since the magnetic behavior of the core is not clear during this transient, one cannot exclude that some steady state may be reached with a minor loop floating midway. In all cases observed, however, the steady state is reached directly on the point of the positive branch of the characteristic. In view of the general steepness of this branch, the jump at $N_cI_c^{**}$ is much more conspicuous in

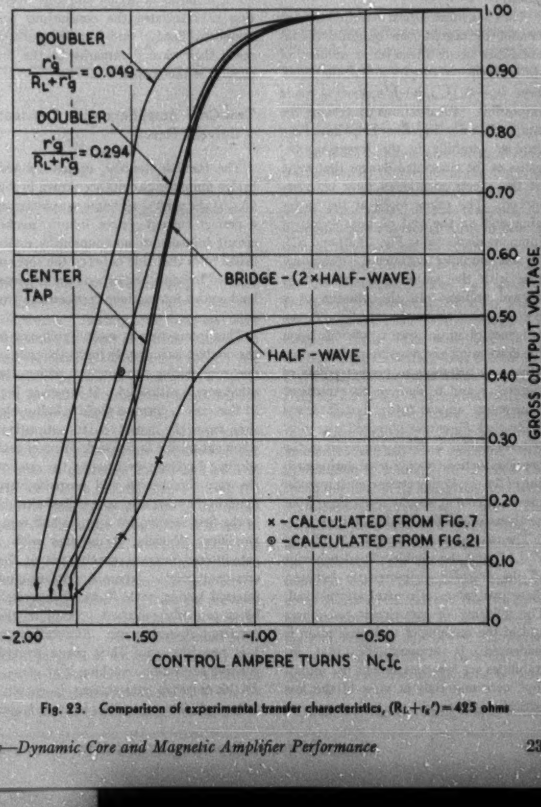


Fig. 23. Comparison of experimental transfer characteristics, $(R_L+r_g')=425$ ohms

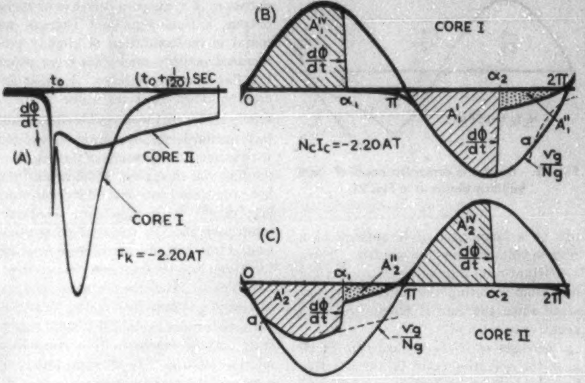


Fig. 24. Reset of a doubler with mismatched cores; normal gate voltage

terms of output. The oscillograms of Fig. 17 show the operation before and after the jump.

The phenomenological description given cannot be better than qualitative; instabilities occur whenever a number of different effects conspire so that the condition $(r_g+R_L)\int_0^T i_g dt+\int_0^T v_\gamma dt=0$ is made impossible. Idealizations concerning the shape of the floating minor loops have been avoided carefully in the foregoing, because of the associated danger that some of the effects mentioned may be overlooked. The minor loops at the beginning and at the end of both types of jump are displayed in Fig. 19(B) and (C), for the amplifier considered in conjunction with the major loop obtained at normal voltage. In the absence of a better quantitative knowledge of the influence of initial core conditions upon the downswing and upswing processes in a cycle, any attempts to predict points of instability and to analyze the transient phenomena appear futile, but it seems likely that F_K curves obtained with procedures similar to those employed in the previous sections may be used during each one of the ratcheting cycles; such reversal curves, of course, should be obtained from pertinent initial flux-level conditions.

The two jumps observed at $|N_cI_c^*|$ and $|N_cI_c^{**}|$ and the double-valued behavior of the transfer characteristic between these two values of control are justified. The influence of gate-circuit resistances and of the amount of underexcitation is evidenced. It appears also that instabilities are less conspicuous in "round-top" core materials in view of the less critical significance of the degree of residual nucleation existing at the beginning of the flux reversal. This description given in terms of actual amplifier operation substantiates the conclusions presented by Batdorf and Johnson and supports their closing remarks to the discussion of that paper.[13]

Two-Core Amplifiers with Constant-Current Reset

The familiar doubler, center-tap, and bridge amplifier circuits are shown in Fig. 20. If the control circuit is supplied from a perfect current source, infinite control circuit impedance, no coupling is established from that side between the cores of the two building blocks, but the common load causes interactions between the two gate circuits.[1]

This interaction is most significant in the doubler as it greatly limits the reset of one core during the time in which the other core is saturated. It is seen in Fig. 21 that core I, during a positive half-cycle, rises from the preset ϕ_p to saturation; when saturation is reached, firing of gate winding I occurs; eventually, the reset of the core flux begins and proceeds very much in the same way as described already in the first section, for a single half-wave amplifier. Namely, it initiates with a gate-limited interval which ends at the crossover point *a*, where a constant-mmf interval begins, with $N_g d\phi_1/dt$ expressed by a properly matched portion of the pertinent F_K-test curve. But while core I is resetting, core II is rising toward positive saturation, reaching it at $\omega t=\alpha_2$. At this time the gate current i_{g2}, practically negligible so far, rises to much larger output values and, accordingly, the "limiter voltage" v_{lim} which up to this time was nearly equal to v_g becomes $v_{lim}=v_g\times r_g'/(r_g'+R_L)$ (where r_g' is the ohmic resistance of one gate winding, corrected for forward rectifier drop). Commonly, this new value of the limiter voltage is, in absolute value, lower than the value of $N_g d\phi_1/dt$ at that time; therefore, the rectifier γ_1 restrikes into conduction and for the rest of the half-cycle the reset of the core proceeds with $N_g d\phi_1/dt=v_g\times r_g'/(r_g'+R_L)$. It appears, hereby, that the amount of flux reset caused by any given N_cI_c is smaller than in the simple half-wave amplifier; i.e., larger values of N_cI_c are needed now to obtain a wanted reset. This is particularly true in respect to the upper part of the transfer characteristic.

The search for the angle of firing α_1, for given N_cI_c, is simple if one assumes that the two building blocks are identical, i.e., have identical windings and identical dynamic behavior of the cores. This assumption shall be removed later. In fact, the half-cyclic volt-time area of core reset $(A^{I}+A^{II})$ in the steady state equals the volt-time area A^{IV} of prefiring upswing from $\omega t=0$ to $\omega t=\alpha_1$; the latter area equals $(A^{I}+A^{III})$. Hence, to obtain α_1 for given N_cI_c, the rest of core I is examined at first with the procedures of the first section, as if it were a simple half-wave amplifier. On the other hand, the curve $v_{lim}=v_g\times r_g'/(r_g'R_L)$ is traced. The vertical line which makes $A^{II}=A^{III}$ determines $\alpha_2=\alpha_1+\pi$ and thus α_1. The condition $A^{II}=A^{III}$ indicates that, with $r_g'>0$, the core II cannot reach saturation during the first interval of gate-limited reset of core I. In the example shown in Fig. 21, core II saturates while core I is resetting in a constant-mmf interval. Saturation of core II could also occur at some later time, at which core I has entered a new gate-limited interval. In some special situations of unusually large gate-winding resistances or constant-mmf resets proceeding with rather small values of $N_g d\phi_1/dt$, the sudden decrease of v_{lim} at the instant of firing of core II may not be sufficient to cause γ_1 to restrike. The constant-mmf process then continues until $N_g d\phi_1/dt=v_g r_g'/(r_g'+R_L)$; the final gate-limited process takes over from here. The contour of the area A^{II} is somewhat different from the one shown in Fig. 21, but this has no relevance for the determination of α_1; the general condition $A^{II}=A^{III}$ still applies.

Evidently, the transfer characteristics of the doubler (i.e., expressing gross- or net-output volt-time areas or also rectified averages of output current versus

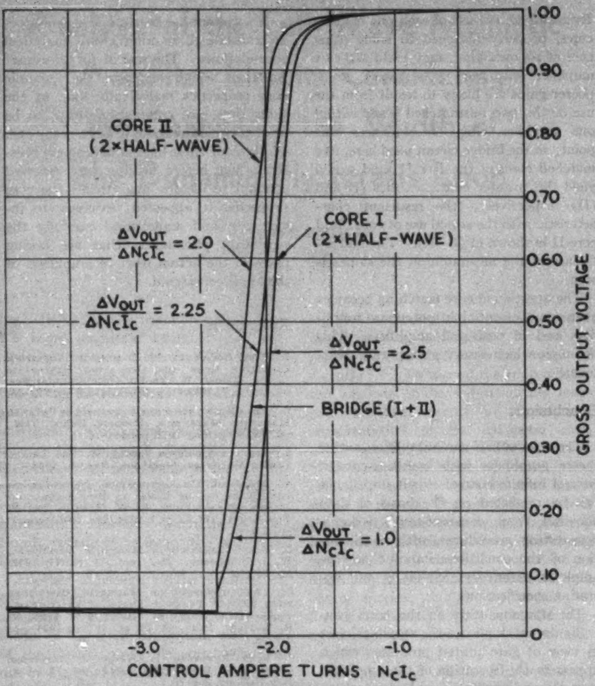


Fig. 25. Influence of core mismatch upon the transfer characteristic of a bridge amplifier, normal gate voltage

N_cI_c) cannot be obtained simply by doubling the corresponding outputs of the half-wave amplifier considered in the first section. Moreover, the half-cyclic volt-time area of gross output here depends not only on the value of the constant-control current I_c, but also on the value of the load resistance R_L. That is, the angle of firing for given N_cI_c varies by varying R_L, simply because R_L appears in the expression of the gate-limiter voltage $v_{ltm}=v_g r_v'/(r_g'+R_L)$ that limits the reset after the firing of the other core. (No such dependency is found in the doubler with zero control circuit impedance; there, the reset takes place with an entirely different mechanism; see reference 6.)

Predictions of behavior are possible also for nonresistive loads; but the procedure is more laborious because the apparent extinction angle β, at which reset initiates, for any given load depends on α, as mentioned already in connection with Fig. 9. Moreover, in the present case, the wave form of i_{g2} after firing (and thus of $v_{ltm}=-r_g^1 i_{g2}$) for any given load depends on α. Both i_{g2} after firing and β are obtainable in terms of α from nomograms and formulas given in reference 12, but since α is not known in advance, a repetitive cut-and-try procedure is needed for any assigned value of N_cI_c.

The interaction between gate circuits is practically less important in the center-tap circuit, as generally known and seen in Fig. 22. Here too, the core I during the positive half-cycle and during an initial part of the reset half-cycle behaves as described for the half-wave amplifier. At $\omega t=\alpha_1+\pi$ core II saturates and the sudden increase of i_{g2} modifies the gate-limiter voltage v_{ltm} which rises from the previous value $v_{ltm}=v_g$ to the value $v_{ltm}=v_g[1+R_L/(r_g'+R_L)]$. The constant-mmf process in progress is not interrupted by this sudden change here. Only the transition to the final gate-limited interval at b is delayed, and the reset during this last interval proceeds with $N_gd\phi_1/dt=v_g[1+R_L/(r_g'+R_L)]$. The over-all volt-time area of reset in the center-tap amplifier therefore is not much different from that of the half-wave amplifier for equal values of N_cI_c. The similarity of transfer characteristics of the half-wave and of the center-tap amplifier, recognized in reference 1, is well justified. The search for the angle of firing again is based on the equality of the areas A^{II} and A^{III}. The influence of R_L upon α_1, for given N_cI_c, is evidently much less significant than in the doubler.

Finally, the interaction between gate circuits is almost negligible in the bridge circuit. Here, after firing of core II, v_{ltm} is modified only by the change of the voltage drop in the rectifier γ_2. The transfer characteristic of this amplifier is found to be even more closely equal to the characteristic of the half-wave amplifier, on a half-cyclic basis.

Fig. 23 shows experimental transfer characteristics for the amplifiers considered in conjunction with transfer characteristics of the building block operating as a half-wave amplifier. For the doubler, two different characteristics are obtained for two different values of the ratio $r_g'/(R_L+r_g')$.

Two-Core Amplifiers with Mismatched Building Blocks

The most common case of mismatch, that caused by differences in the dynamic behavior of the two cores, is considered here. (Extensions of the reasoning to mismatches of number of turns in gate or control windings and to cases of asymmetries of the two half-waves of gate voltage are immediate.) The F_K curves of two different cores are shown in Fig. 24(A) for a common value of $F_K=-2.20$ AT (ampere turn). The operation of a doubler using those two cores is shown in Fig. 24(B) for the same value of control $N_cI_c=-2.20$ AT. The reset of each core, before firing of the other component, takes place as if each building block were an independent half-wave amplifier. Also information is available to describe the remaining part of its half-cyclic reset with $N_gd\phi_1/dt=v_g\times r_g'/(r_g'+R_L)$, etc. The search for the unknown angles of firing α_1 and α_2 is somewhat more laborious than in the previous case of a doubler with matched cores, because the volt-time areas of half-cyclic reset of the two cores here are not equal, that is $\phi_{s1}-\phi_{p1}\neq\phi_{s2}-\phi_{p2}$ and therefore $\alpha_2\neq\alpha_1+\pi$. However, the prefiring area A^{IV} of each core still equals its own half-cyclic area of reset $(A^{I}+A^{II})$.

Therefore, the two firing angles α_1 and

α_2 must be such as to satisfy simultaneously the two relations:

1. Prefiring area A_1^{IV} = half-cyclic area of reset ($A_1^{I}+A_1^{II}$), for core *I*.
2. Prefiring area A_2^{IV} = half-cyclic area of reset ($A_2^{I}+A_2^{II}$), for core *II*.

The cut-and-try search for α_1 and α_2 at given N_cI_c can be systematized as follows: for each core $d\phi/dt$ is plotted as a function of time over the whole reset half-cycle, both as if the core were operating in a half-wave amplifier and as if it were limited to the values $d\phi/dt = v_{1lm}/N_g$. Then, tentative values are assigned to α_1, and α_2 is determined accordingly so as to satisfy the first of the two previous conditions. Likewise, tentative values are assigned to α_2, and α_1 is determined accordingly so as to satisfy the second of these two conditions. This yields the correct values of α_1 and α_2 which satisfy simultaneously both conditions. (Note that in cases of core mismatch, of course, firing of one core can well occur while the other is still in its first gate-limited interval.)

Since the two building blocks are much less interdependent in the case of the center-tap amplifier, and even more so in the case of the bridge amplifier, the cyclic output of these two amplifiers is more nearly equal to the sum of the half-cyclic outputs of the two mismatched components in independent half-wave operation.

It would seem from all this that a matching of the cores would not be an indispensable prerequisite in the production of the 2-core amplifiers considered. After all, the resulting transfer characteristic obtained with the use of cores of two different kinds lies somewhere between the characteristics that one would obtain with the use of matched cores of the one or of the other kind. Unfortunately, in most amplifier applications the quantity of primary interest is the slope of the transfer characteristic at any given output; i.e., incremental gains are the real matter of concern. Even though the use of each one of these cores, properly matched to some other core of its own kind, may yield a 2-core amplifier with very good gains, much poorer gains are likely to result from the use of the two mismatched cores within one amplifier. Fig. 25 evidences this point; in the bridge circuit used here, two matched cores of the I or II kind would yield the transfer characteristics (I) and (II) respectively; the resultant characteristic with the actual use of core I and core II is shown in (I+II). The bad influence of the mismatch on the slopes is seen.

The question of core matching becomes even more essential in polyphase amplifiers and in push-pull amplifiers. This is a topic which deserves a separate treatment.

Conclusions

The reset of half-wave amplifiers and of 2-core amplifiers with constant-current control infinite control circuit impedance can be predicted on the basis of data acquired from constant-mmf dynamic core-testing procedures, with consideration of the conditions imposed by the amplifier parameters, circuitry, and operating specifications.

The dynamic tests on the cores generally do not duplicate the amplifier reset in view of gate-limited processes which appear in the operation of the amplifier. But, the information supplied by those tests is most valuable and adequate if used properly with criteria that have been described in the text. These criteria have been derived from the assumption that the dynamic behavior of a grain-oriented 50% nickel-iron core resetting from pronounced saturation is characterized by the existence of a unique relationship among instantaneous core flux, core magnetomotive force, and rate of change of core flux. The many examples support the validity of this assumption and, in fact, confer a broad significance to the results of those core-testing methods.

No attempt has been made in this paper to translate those criteria into analytical formulations. The search for core-reset functions which represent the dynamic core properties realistically and, at the same time, are reasonably simple to be used in circuit analyses is not an easy one. Work along these lines is in progress presently, but before making any renewed attempt to idealize analytically the core properties it appeared necessary to investigate and understand carefully the significant phenomena that are taking place in the actual reset of amplifiers of the kinds considered.

References

1. The Influence of Magnetic Amplifier Circuitry Upon the Operating Hysteresis Loops, **Harold W. Lord.** *AIEE Transactions,* vol. 72, pt. I, 1953 (Jan. 1954 section), pp. 721–28.

2. Flux Resetting Characteristics of Several Magnetic Materials, **Hoobert Huhta.** *Ibid.,* vol. 73, pt. I, May 1954, pp. 111–14.

3. The Transductor (book), **A. Uno Lamm.** Esselte Aktiebolag, Stockholm, Sweden, 1948.

4. Magnetic Characteristics Pertinent to the Operation of Cores in Self-Saturating Magnetic Amplifiers, **R. W. Roberts.** *AIEE Transactions,* vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 682–90.

5. Self-Saturation in Magnetic Amplifiers, **W. J. Dornhoefer.** *Ibid.,* vol. 68, pt. II, 1949, pp. 835–50.

6. The Operation of Magnetic Amplifiers with Various Types of Loads—Part II—Controlling the Angle of Firing, **L. A. Finzi, R. R. Jackson.** *Ibid.,* vol. 73, pt. I, July 1954, pp. 279–88.

7. Flux Reversal in Magnetic Amplifier Cores, **F. J. Friedlaender.** *Ibid.,* vol. 75, pt. I, July 1956, pp. 268–78.

8. Dynamic Hysteresis Loop Measuring Equipment, **H. W. Lord.** *AIEE Transactions (Electrical Engineering),* vol. 71, pt. I, Sept. 1952, pp. 269–72.

9. An Accurate Electronic Tracer for Dynamic Characteristics of Magnetic Materials, **Emil Kittl.** *AIEE Transactions,* vol. 74, pt. I, Sept. 1955, pp. 407–18.

10. Flux Reset in Magnetic Amplifier Cores, **B. Seddon.** National Electronics Conference, Chicago, Ill., Oct. 3, 1955.

11. The Operation of Magnetic Amplifiers with Various Types of Load—Part I—Load Currents for Given Angle of Firing, **L. A. Finzi, R. R. Jackson.** *AIEE Transactions,* vol. 73, pt. I, July 1954, pp. 270–78.

12. An Instability of Self-Saturating Magnetic Amplifiers Using Rectangular Loop Core Materials, **S. B. Batdorf, W. N. Johnson.** *Ibid.,* vol. 72, pt. I, July 1953, pp. 223–28.

Analysis of the Full-Wave Magnetic Amplifier Circuits Considering the Change of the Width of the Dynamic Hysteresis Loop

TADASHI KIKUCHI
ASSOCIATE MEMBER AIEE

RECENTLY the behavior of the full-wave magnetic amplifier was analyzed in the case of the very low control circuit resistance and the correlation between the instantaneous hysteresis loop and the control characteristics was demonstrated.[1] Concerning the transient behavior, the approach described in the literature[1,2] permits computation of an equivalent time constant. This concept may be applied to the case of the "slow transient." However, when the control circuit resistance is large, the approach loses its meaning.

Another approach to the solution of the problem was proposed recently by adopting difference equations to the model of the feedback system composed of the time-delaying element,[3] which is one of the chief features of the magnetic amplifier. Though the method is mathematically elegant, it cannot explain the correlation between the core characteristics and the amplification, because of the assumption of the idealized core characteristics.

Among the several approaches to the problem, the most suitable seems to be the one that discusses the change of the modes of the circuit.[4] This method enables, first, a discussion of the phenomena in detail during a half-cycle and, following that, a description of the relations of the succeeding half-cycles. Several papers following this approach appeared recently.[5–8] To extend the approach for the analysis of the self-saturating magnetic amplifier circuit, an appropriate representation of the magnetic core, which greatly influences the operation of the amplifier, must be found. One can easily recognize that the mere idealization of the core characteristics, such as the conventional rectangular hysteresis loop, is not sufficient to solve the circuit equation. It is necessary to adopt an appropriate model to represent the core characteristics, though the actual behavior of the magnetic core in the magnetic-amplifier circuit is very complicated.[9,10]

The approach utilizing the equivalent core resistance to analyze the mechanism of the self-saturating magnetic amplifier was reported recently,[8] and presented a reasonable result, though the case treated is rather limited and simplified.

As the first approximation to describe the core characteristics, the author proposed equations 1 and 1(A) recently[11,12] which take into account the change of the dynamic hysteresis loop. This representation seems to be convenient for the solution of the problem as stated here.

In order to analyze the full-wave magnetic amplifier, considering the change of width of the dynamic hysteresis loop, it is obvious that the conventional method of analysis cannot be used. Therefore the approach using the difference equation will be discussed in more detail from a rather new standpoint, which lends itself to the complete solution, transient and steady state, of the full-wave magnetic amplifier.

In order to clarify the concept of the time constant in the discrete system, the correspondence between the systems represented by the linear differential equation and the one represented by the linear difference equation, both of the first order, is discussed. Then the difference equation is derived to represent the behavior of the magnetic amplifier. The meaning of frequency response method to measure the time constant[2,11] of the discrete system is also made clear. The results can explain several features of the full-wave magnetic amplifier from the complete solution viewpoint.

Nomenclature

A_{v0} = voltage amplification, on common turns basis, of basic full-wave magnetic amplifier
a = coefficient of differential equation 10
b = constant of differential equation 10
C = amplitude of periodic solution (equation 24)
C_2 = arbitrary constant
c_1 = arbitrary constant
E_c = control voltage, direct current, volts
E_{ce} = effective d-c control voltage, volts (equation 56)
E = a-c supply voltage, volts, Maximum (equation 63)
E' = resultant reverse voltage, volts, maximum (equation 53)
e_{ac} = a-c supply voltage, volts, instantaneous (equation 63)
e_{ac}' = resultant reverse voltage, volts, instantaneous (equation 53)
f = supply frequency, cycles per seconds
I_c = d-c control current, amperes (equation 40)
I_{c0} = control current corresponding to average value of dynamic hysteresis loop, amperes (equation 7)
I_{00} = control current corresponding to static coercive force, amperes (equation 1)
i = current of windings, amperes (equation 1)
i_c = control current, instantaneous, amperes
i_1 = gate current of mesh 1, amperes
i_2 = gate current of mesh 2, amperes
k = coefficient representing conductance of core to eddy current, mhos (equation 1)
N = number of turns
n = integer representing number of half-cycles
R_c = resistance of control circuit, ohms
R_G = resistance of gate circuit, ohms
R_L = load resistance, ohms
R_{LG} = sum of R_G and R_L, ohms (equation 59)
R_{LG}' = equivalent resultant resistance, ohms (equation 52)
R_e = equivalent effective core resistance, ohms (equation 4)
T = period of supply voltage, seconds
T_r = time constant, seconds (equation 46)
T_d = dead-time lag, seconds (equation 51)
Δt = time interval explained by Fig. 7, seconds (equation 50)
Δt_0 = additional dead-time lag, seconds (equation 51)
t = time, seconds
t_n = discrete instant of equal interval, seconds
$V_L(n)$ = average output voltage during nth half-cycle, volts (equation 31)
$\Delta V_L(n)$ = difference between maximum output voltage and $V_L(n)$, volts (equation 32)
ΔV_{L0} = steady-state value of $\Delta V_L(n)$ in case of basic full-wave magnetic amplifier circuit, volts (equation 39)

Paper 57-97, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 24, 1956; made available for printing November 21, 1956.

TADASHI KIKUCHI is with the Microwave Research Institute, Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

The author wishes to thank Professors Ernst Weber and E. J. Smith, Polytechnic Institute of Brooklyn, for their kindness and encouragement and A. Aronowitz, Polytechnic Institute of Brooklyn, for his kind comments and his help in improving the English expression of the paper. The author also wishes to thank Professors Y. Watanabe and K. Fukushima of Tohoku University for their kind encouragement and guidance.

The research reported in this paper has been made possible, in part, through the sponsorship of the Office of Naval Research under contract *Nonr-839(05)*.

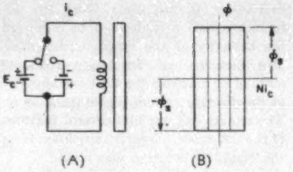


Fig. 1. Dynamic hysteresis loop for rectangular voltage

V_m = average value of maximum output voltage including internal resistance drop, volts (equation 8)
X_n = dependent variable of difference equation (equation 12)
x = dependent variable of differential equation (equation 10)
α = coefficient of difference equation (equation 12)
α_n = firing angle of nth half-cycle (equation 62)
β = constant of difference equation 12
γ = sinusoidal driving force, maximum value (equation 23)
δ = constant defined by equation 26
ϵ = base of natural logarithm (equation 20)
θ_0 = phase angle of transition from mode 0 to mode 1 (equation 86)
θ_1 = phase angle of transition from mode 1 to mode 2 (equation 84)
θ_2 = phase angle of transition from mode 3 to mode 0 (equation 68)
$\Delta\theta_2$ = difference between π and θ_2 (equation 71)
$\sigma(\theta_1)$ = coefficient of commutation due to transition angle θ_1 (equation 78)
$\sigma(\Delta\theta_2)$ = coefficient of commutation due to transition angle θ_2 (equation 78)
ρ = coefficient of commutation due to circuit resistances in mode 1 (equation 73)
ρ' = coefficient of commutation due to circuit resistances in mode 0 (equation 74)
Φ_s = saturation value of core flux, webers (equation 1)
Φ = instantaneous core flux, webers (equation 1)
Φ_1 = instantaneous value of flux in core 1, webers (equation 33)
Φ_2 = instantaneous value of flux in core 2, webers (equation 33)
$\Delta\Phi(n)$ = amount of reset flux during nth half-cycle, webers (equation 37)
$\Delta\Phi(\infty)$ = amount of reset flux during half-cycle of steady state, webers (equation 38)
$\Delta^2\Phi(\theta_1)$ = decrease of amount of reset flux during mode 1 by commutation, webers (equation 69)
$\Delta^2\Phi(\Delta\theta_2)$ = decrease of amount of reset flux during mode 0 by commutation, webers (equation 70)
$\Delta^2\Phi$ = resultant decrease of amount of reset flux due to commutation, webers (equation 75)
$\Delta'\Phi(n)$ = amount of reset flux before firing of core in case of abnormal modes, webers (equation 90)
$\Delta\Phi''(n)$ = amount of reset flux after firing of core in case of abnormal modes, webers (equation 91)
ω = angular frequency of a-c supply voltage, radians per second (equation 63)

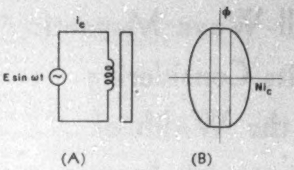


Fig. 2. Dynamic hysteresis loop for sinusoidal voltage

ω = angular frequency of sinusoidal input to discrete system, radians per second (equation 23)

Preliminary Considerations

Assumption and Approximate Representation of the Core Characteristic

To simplify the mathematical expressions, consider the circuit phenomena on the common-turns basis, i.e., that the numbers of turns of the saturating reactor are all equal to N. Then the characteristics of the saturating reactor core with the rectangular hysteresis loop may be approximately represented as follows:

(a) for the descending branch of the dynamic hysteresis loop

$$N\Sigma i = NI_{00} - kd\Phi/dt$$

and

$$|\Phi| < \Phi_s \text{ and } d\Phi/dt < 0 \tag{1}$$

(b) for the ascending branch of the dynamic hysteresis loop

$$N\Sigma i = -NI_{00} - kd\Phi/dt$$

and

$$|\Phi| < \Phi_s \text{ and } d\Phi/dt > 0 \tag{1A}$$

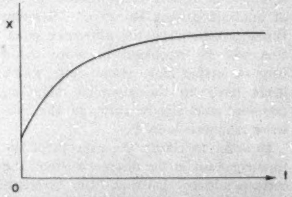


Fig. 3. Solution of first-order differential equation

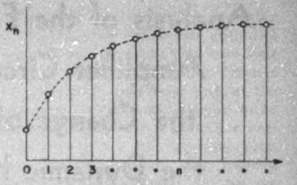


Fig. 4. Solution of first-order difference equation

In the foregoing equation, Φ and Φ_s are, respectively, the instantaneous value of the magnetic flux of the core and its saturation value; k and NI_{00} are the constants of the core; and $N\Sigma i$ means the resultant instantaneous ampere-turns applied to the magnetic core. The positive direction of the ampere-turns is assumed for the resetting action.

Physical Meaning of the Constant "k"

First to be considered is the case shown in Fig. 1, where a rectangular voltage is applied to the saturating reactor, to excite it along the dynamic hysteresis loop of Fig. 1(B). Then neglecting the circuit resistance, the control voltage E_c is directly related to the induced electromotive force of the core, that is,

$$E_c = Nd\Phi/dt \tag{2}$$

Substituting this into equation 1, the following relation evolves:

$$1/k = E_c/N^2(i_c - I_{00}) \tag{3}$$

If the expression

$$R_0 = E_c/(i_c - I_{00}) \tag{4}$$

is adopted, the constant k can be represented as

$$1/k = R_0/N^2 \tag{5}$$

Therefore k can be considered as the conductance of the magnetic core as seen by the eddy current.

Anticipating a later discussion, consider the case of the so-called "sine flux" excitation, i.e., the saturating reactor is excited by an ideal sinusoidal voltage source, E sin ωt. Then, during the resetting half-cycle of the sinusoidal voltage

$$i_c = I_{00} + (k/N^2)E \sin\omega t \tag{6}$$

and its average value during the half-cycle, I_{c0}, is

$$I_{c0} - I_{00} = \left(\frac{2}{T}\right)\left(\frac{k}{N^2}\right)E\int_0^{T/2} \sin \omega t dt \tag{7}$$

On the other hand, the maximum output voltage of the magnetic amplifier including the gate circuit impedance drop, V_m, can be defined as the equation

$$V_m=\frac{2}{T}E\int_0^{T/2}\sin\omega t dt \tag{8}$$

Therefore, k can be represented as follows:

$$\frac{k}{N^2}=\frac{I_{c0}-I_{00}}{V_m} \tag{9}$$

The dynamic behavior of the magnetic core under this condition can be represented as shown in Fig. 2(B).

Correspondence of the Continuous and Discrete Systems

Since this paper will analyze the self-saturating magnetic amplifier by means of difference equations, it is necessary at this point to discuss the analogy between the solution of the first-order linear differential and difference equations. In particular, the concept of an equivalent time constant will be extended to the difference equation solution.

Consider

$$dx/dt+ax=b \tag{10}$$

where a and b are constants and the general solution of the equation is, representing the arbitrary constant by C_1,

$$x=C_1\epsilon^{-at}+b/a \tag{11}$$

which can be shown, e.g., in Fig. 3.

On the other hand, the first-order linear difference equation which describes the behavior of the simpler discrete system as found in the literature,[4-8] can be represented as

$$X_{n+1}-\alpha X_n=\beta \tag{12}$$

when n is any arbitrary integer. The general solution of the foregoing equation is

$$X_n=C_2\alpha^n+\frac{\beta}{1-\alpha} \tag{13}$$

where α and β are given constants, and C_2 is an arbitrary constant defined at the discrete point n. The solution X_n can be represented as shown in Fig. 4. In the case of the magnetic amplifier, certain physical quantities of every half-cycle such as the reset flux level, the firing angle, or the average output current, etc., will be represented by the quantity X_n.

In order that these two systems have stable solutions, it is necessary respectively that

$$a>0 \tag{14}$$

and

$$|\alpha|<1 \tag{14A}$$

then, the steady-state solution is, as is well known, the equation following

$$\lim_{t\to\infty} x=\frac{b}{a} \tag{15}$$

and

$$\lim_{n\to\infty} X_n=\frac{\beta}{1-\alpha} \tag{15A}$$

Initial conditions for both systems are

$$x|_{t=0}=c_1+b/a \tag{16}$$

$$X_n|_{n=0}=C_2+\beta/(1-\alpha) \tag{16A}$$

The necessary and sufficient condition for x and X_n to coincide at the discrete instant t_n,

$$t_n=\frac{T}{2}xn\ (n=1, 2, 3, \dots) \tag{17}$$

can be easily found as

$$c_1=C_2 \tag{18}$$

$$b/a=\beta/(1-\alpha) \tag{19}$$

$$\alpha=\epsilon^{-a(T/2)} \tag{20}$$

where $T/2$ is the duration of the half-cycle which is also the interval between two neighboring discrete points.

The time constant of the continuous system represented by equation 10 is

$$T_r=1/a \tag{21}$$

Therefore, the time constant of the discrete system can be defined as follows, using the foregoing relation of correspondence:

$$T_r=\left(\frac{T}{2}\right)\left(\frac{1}{\ln(1/\alpha)}\right)=\left(\frac{1}{2f}\right)\left(\frac{1}{\ln(1/\alpha)}\right) \tag{22}$$

where f is the a-c supply frequency. Similar form of the time constant was given by Gardiner[14] from a somewhat different approach. Thus, when the behavior of a magnetic amplifier can be described by a first-order linear difference equation, its steady-state value and transient time constant can be easily obtained without any detailed calculations.

Frequency Response Method in Discrete System

To determine the dynamic behavior of the magnetic amplifier, measurements by the frequency response method are used. From the consideration, the meaning of the frequency response method as applied to the discrete system will be checked.

Assume that the following difference equation is satisfied for a discrete system under the frequency response method

$$X_{n+1}-\alpha X_n=\gamma\epsilon^{jn\omega_0/2f} \tag{23}$$

where ω_0 is the angular frequency of the driving force and γ is a given constant. Assume the steady-state solution of equation 23 as follows:

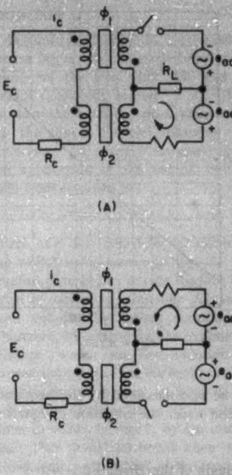


Fig. 5. Basic modes of full-wave magnetic amplifier

$$X_n=C\epsilon^{jn\omega_0/2f} \tag{24}$$

Substituting this in the previous equation, the following relation is obtained:

$$\epsilon^{j\omega_0/2f}-\alpha=\gamma/C \tag{25}$$

Putting

$$\alpha=\epsilon^{-\delta/2f} \tag{26}$$

and substituting this in equation 25

$$\epsilon^{j\omega_0/2f}-\epsilon^{-\delta/2f}=\gamma/C \tag{27}$$

Assuming

$$\omega_0/2f<<1 \text{ and } \delta/2f<<1 \tag{28}$$

and expanding equation 27, the following relation is obtained:

$$\frac{\delta}{2f}\left(1+\frac{j\omega_0}{\delta}\right)\cong\frac{\gamma}{C} \tag{29}$$

In the case

$$\omega_0=\delta \tag{30}$$

a phase shift of 45 degrees will be observed between the input and output of the system. On the other hand, from equations 20 and 26, the quantity "δ" will be found to be identical with "a" of the corresponding differential system, which is the inverse of the time constant T_r.

Therefore, the time constant of the discrete system defined in equation 22

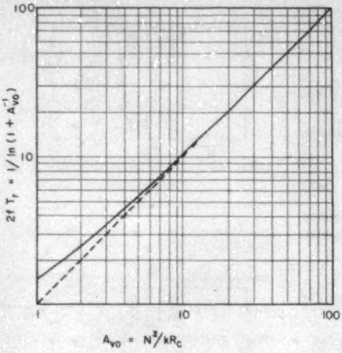


Fig. 6. Relation between voltage amplification and time constant of full-wave magnetic amplifier

can be also measured by the frequency method under the condition of equation 28.

Analysis of the Basic Full-Wave Magnetic Amplifier Circuit

The basic operation of the full-wave magnetic amplifier may be manifested by the sequence of the two modes shown in Fig. 5(A) and 5(B) alternately repeated every half-cycle. This representation, although neglecting the effect of commutation of the gate circuit, completely satisfies the transient case when the control voltage disappears suddenly in the full wave or doubler circuit with resistance load. The effect of the commutation for the full-wave amplifier will be discussed in the next section.

If the reset flux during the nth half-cycle of the a-c supply frequency is represented as $\Delta\Phi(n)$, then the average output voltage during the $(n+1)$ half-cycle, including the resistance drop of the gate circuit, is as follows:

$$V_L(n+1) = V_m - 2fN\Delta\Phi(n) \tag{31}$$

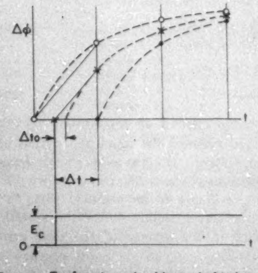


Fig. 7. Explanation of additional dead-time lag

where V_m is defined by equation 8, or

$$\Delta V_L(n+1) = 2fN\Delta\Phi(n) \tag{32}$$

where

$$\Delta V_L(n+1) = V_m - V_L(n+1) \tag{32A}$$

The expression $\Delta\Phi(n)$ is convenient for the discussion of the theory of the magnetic amplifier.

To seek the relation satisfied by $\Delta\Phi(n)$, the equation of the control circuit is solved by using the expression for the core characteristics discussed in the previous section. That is,

$$R_c i_c - N(d\Phi_1/dt + d\Phi_2/dt) = E_c \tag{33}$$

To solve equation 33 for a resetting half-cycle of core 1 shown in Fig. 5(A), the relation of equation 1 can be used as follows:

$$i_c = I_{00} - \frac{k}{N}\,\frac{d\Phi_1}{dt} \tag{34}$$

Then, from the equations, 33 and 34, the following relation is deduced:

$$-(N + kR_c/N)d\Phi_1/dt = (E_c - R_cI_{00}) + Nd\Phi_2/dt \tag{35}$$

Assuming that the foregoing equation is satisfied during the nth half-cycle of the a-c source voltage, then during that half-cycle, the flux of core 1 is reset and that of core 2 is driven to saturation in its gating half-cycle. Therefore, putting

$$\left[\int_0^{T/2} |d\Phi_1/dt|\,dt\right] \text{resetting } n\text{th half-cycle} = \Delta\Phi(n) \tag{36}$$

and

$$\left[\int_0^{T/2} |d\Phi_2/dt|\,dt\right] \text{gating } n\text{th half-cycle} = \left[\int_0^{T/2} |d\Phi_2/dt|\,dt\right] \text{resetting } (n-1)\text{th half-cycle} = \Delta\Phi(n-1) \tag{36A}$$

the following difference equation is derived from equation 35:

$$N(1 + kR_c/N^2)\Delta\Phi(n) = (E_c - R_cI_{00})/2f + N\Delta\Phi(n-1) \tag{37}$$

It will be easily seen that equation 37 cannot have a stable solution if the coefficient k is zero, that is, if there is no dynamic loss of the core.

STEADY-STATE CHARACTERISTICS

The steady-state value of the reset flux, $\Delta\Phi(\infty)$, is

$$\Delta\Phi(\infty) = \frac{E_c - R_cI_{00}}{kR_c/N} \times \frac{1}{2f} \tag{38}$$

Therefore, the steady-state output voltage is, from equation 32,

$$\Delta V_{L0} = \frac{N^2}{kR_c}(E_c - R_cI_{00}) \tag{39}$$

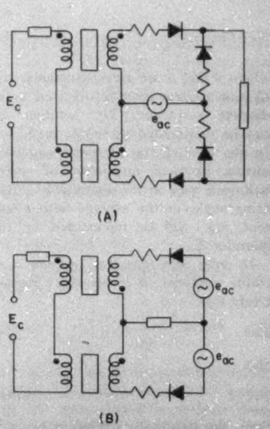
Putting

$$I_c = E_c/R_c \tag{40}$$

and using the expression of equation 9, equation 40 can be represented as follows,

$$\frac{\Delta V_{L0}}{V_m} = \frac{I_c - I_{00}}{I_\infty - I_{00}} \tag{41}$$

This expression coincides with the concept that the full-wave amplifier can be considered as a current-controlled device. But it also indicates that the slope of the control characteristics of the amplifier has a close relation to the width of the dynamic hysteresis loop. Thus the adaptation of equation 1, for the dynamic characteristics of the magnetic core is sufficient as a first approximation.



(A)

(B)

Fig. 8. Full-wave magnetic amplifier circuits

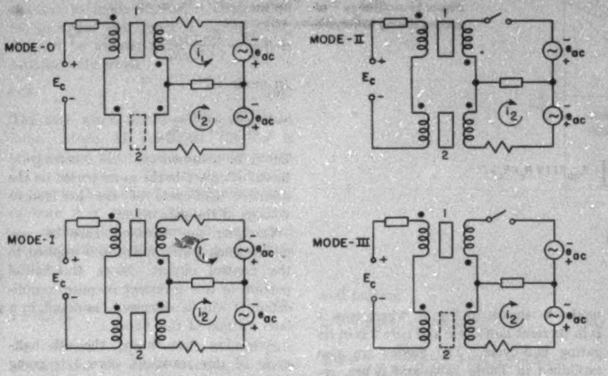


Fig. 9. Modes of full-wave magnetic amplifier circuit affected by commutation

From equations 39 and 5, another way of representing the control characteristics is

$$\frac{\Delta V_{L0}}{E_c - R_c I_{00}} = \frac{R_0}{R_c} = \frac{N^2}{kR_c} \qquad (42)$$

Expressing the voltage amplification by the equation

$$A_{v0} = \frac{\Delta V_{L0}}{E_c - R_c I_{00}} \qquad (43)$$

the following is obtained:

$$A_{v0} = \frac{R_0}{R_c} = \frac{N^2}{kR_c} \qquad (44)$$

and

$$\Delta V_{L0} = R_0(I_c - I_{00}) \qquad (45)$$

That is, the voltage amplification of the magnetic amplifier is proportional to the core resistance of the eddy current, and inversely proportional to the control circuit resistance R_c.

TRANSIENT CHARACTERISTICS

The transient characteristics of the basic full-wave magnetic amplifier is represented by a time constant and dead time. Even in the case of normal excitation, there will be a lag of one-half cycle between the amount of reset flux, $\Delta\Phi(n)$, and the average output voltage $VL(n+1)$, as the relation of equation 31 shows. The time constant is the same for these two quantities, and it can be easily found from equation 37.

Table I. Modes of Half-Wave Magnetic Amplifiers

	Modes 0	I	II	III
$d\Phi_1/dt$	<0	<0	<0	<0
i_1	>0	>0	0	0
$d\Phi_2/dt$	0	>0	>0	0
i_2	>0	>0	>0	>0

TIME CONSTANT

Comparing equations 12 and 37, and using the relation of equation 22, the time constant for this case is given as follows:

$$T_r = \frac{1}{2f}\,\frac{1}{\ln(1+kR_c/N^2)} \qquad (46)$$

which, together with equation 44, gives

$$T_r = \frac{1}{2f}\,\frac{1}{\ln(1+1/A_{v0})} \qquad (47)$$

In Fig. 6, the relation of equation 47 is plotted. In the case of

$$A_{v0} >> 1 \qquad (48)$$

equation 48 is reduced to the well-known relation

$$T_r = \frac{1}{2f} A_{v0} \qquad (49)$$

Fig. 6 shows the well-known tendency of the time constant when the voltage amplification is decreased by increasing the control circuit resistance R_c, as shown in equation 44; the measured value of the time constant usually deviates to the upper side of the straight dotted line, of Fig. 6, which represents equation 49. Therefore this consideration, though neglecting the effect of commutation, yields a better account of the phenomenon than the conventional simpler approach.

DEAD TIME

As shown in equation 32, there exists at least a lag of one-half cycle between the change of the reset flux, $\Delta\Phi$, and the change of output voltage, ΔV_L. Here consider the transient phenomenon during the first half-cycle, assuming that both cores are entirely saturated before the application of the control voltage; then from equation 35:

$$N(1+kR_c/N^2)\Delta\Phi(1) = (E_c - R_c I_{00})x\Delta t \qquad (50)$$

where Δt is determined by the instant of the switch-in as shown in Fig. 7. The amount of reset flux during the first half-cycle is only proportional to Δt, and it does not usually satisfy equation 37. Therefore, by extending the exponential relation, which satisfies equation 37, to seek the point of the curve that crosses the abscissa, it is seen, in general, that this point deviates from the switch-in time, as shown in Fig. 7, by the interval Δt_0. This should be noted when measuring the time constant of the magnetic amplifier with comparatively large control resistance and small time constant. Therefore, the dead time, T_d, of the basic full-wave magnetic amplifier can be thought as follows for the input of step function

$$T_d = \frac{T}{2} + \Delta t_0 \qquad (51)$$

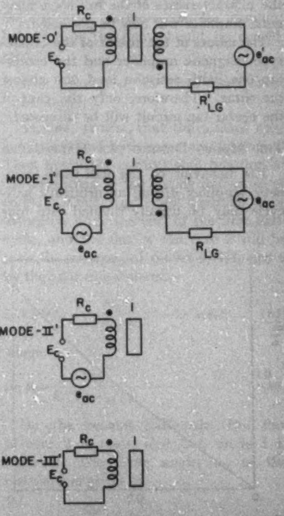


Fig. 10. Equivalent circuits of the modes of full-wave magnetic amplifier

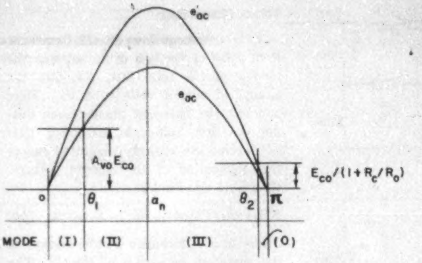


Fig. 11. Change of modes during half-cycle

Full-Wave Magnetic Amplifier Affected by Commutation

In the case of the actual full-wave magnetic amplifier, the modes of the circuit will be somewhat different from those which have been explained, because of the so-called commutation in the gate circuits. Consider the bridge-type or the center-tap magnetic amplifier shown in Fig. 8. During some portion of the resetting half-cycle of the saturable reactor core, the gate circuit rectifier cannot block the gate circuit current, because the instantaneous value of the a-c supply voltage is smaller than the induced electromotive force of the core. The resetting of the core flux will be retarded by this commutation; its effect on the characteristics of the full-wave magnetic amplifier will now be discussed.

The modes of the circuit of the bridge-type magnetic amplifier and the center-tap one, with resistive load, are almost the same. Therefore, only the case of the center-tap circuit will be discussed.

THE MODES DURING THE HALF-CYCLE

The behavior of the half-wave magnetic amplifier during an arbitrary half-cycle may be usually divided into four modes as shown in Fig. 9, where core 1 is in its reset half-cycle and core 2 is in its gating half-cycle. The modes are also explained in Table I; i_1 and i_2 are, respectively, the gate circuit current of the saturable reactor 1 and 2.

If sum of the gate circuit resistance, the gate winding resistance, and the forward resistance of the rectifier are represented by R_G, and the load resistance by R_L, then, as far as the behavior of the saturating reactor 1 is concerned, modes 0, II, and III of Fig. 9 are equivalent to the modes 0′, II′, and III′ of Fig. 10. In Fig. 10

$$R_{LG}'=R_G\left(1+\frac{R_L}{R_L+R_G}\right) \tag{52}$$

$$\begin{aligned} e_{ac}'&=e_{ac}\left(1+\frac{R_L}{R_L+R_G}\right)\\ &=E'\sin\theta \end{aligned} \tag{53}$$

Although mode I is rather complicated, it can be replaced by mode I′ of Fig. 10 to simplify the description. (Further discussion is given in the appendix.)

Using the representation of equation 1 for the core characteristics, the basic relations satisfied in each mode are as follows. In mode 0′

$$-N\frac{d\Phi}{dt}=\frac{E_{c0}+e_{ac}'(R_c/R_{LG}')}{1+R_c/R_{LG}'+R_c/R_0} \tag{54}$$

and

$$i_1=\frac{E_{c0}-e_{ac}'(1+R_c/R_0)}{R_c+R_{LG}'+R_{LG}'R_c/R_0} \tag{55}$$

where

$$E_{c0}=E_c-R_cI_{00} \tag{56}$$

In mode I′

$$-N\frac{d\Phi}{dt}=\frac{E_c+e_{ac}(1+R_c/R_{LG})}{1+R_c/R_0+R_c/R_{LG}} \tag{57}$$

$$i_1=\frac{E_{c0}-e_{ac}(R_c/R_0)}{R_c+R_{LG}+R_{LG}R_c/R_0} \tag{58}$$

where

$$R_{LG}=R_L+R_G \tag{59}$$

In mode II′

$$-N\frac{d\Phi}{dt}=\frac{E_{c0}+e_{ac}}{1+R_c/R_0} \tag{60}$$

In mode III′

$$-N\frac{d\Phi}{dt}=\frac{E_{c0}}{1+R_c/R_0} \tag{61}$$

Let it be assumed that the resetting interval of core 1 is the same period as the negative half-cycle of the a-c source voltage of the gate circuit.

Consider the transient case of amplifier when a step function is applied to the control circuit. Since the initial portion of this transient is quite complicated, it will be discussed, in detail, in a later section of this paper.

Assuming that during the nth half-cycle of this transient, core 1 is going from its saturation to its reset flux level, and that core 2 is heading from its reset to its saturation flux level after having been reset by the amount $\Delta\Phi(n-1)$ in the previous half-cycle, then the firing angle α_n of core 2 is determined by the following relation:

$$\int_0^{\alpha_n} e_{ac}d\theta=N\omega\Delta\Phi(n-1) \tag{62}$$

and the a-c source voltage

$$e_{ac}=E\sin\theta=E\sin\omega t \tag{63}$$

At first, mode I is realized because core 2 is in its exciting interval. The phase angle, θ_1, is defined by the following

$$\begin{aligned} \sin\theta_1&=\left(\frac{E_{c0}}{E}\right)\left(\frac{R_0}{R_c}\right)\\ &=A_{v0}\left(\frac{E_{c0}}{E}\right) \end{aligned} \tag{64}$$

Then, if

$$\theta_1<\alpha_n \tag{65}$$

mode I will continue until the phase

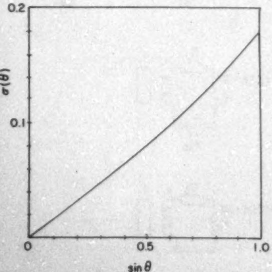


Fig. 12. Coefficient of commutation

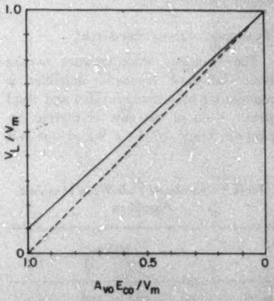


Fig. 13. Influence of commutation to control characteristics

angle θ_1 is reached. The mode then changes to mode II, for, from equation 58, in order that mode I should exist, i_1 must be greater than zero, which is satisfied only when

$$\theta<\theta_1 \tag{66}$$

The case when $\alpha_n<\theta_1$ will be discussed later. Mode II continues till α_n is reached and after core 2 saturates, mode III is in effect. This mode will exist as long as the induced electromotive force of core 2 is smaller than the voltage e_{ac}, defined by equation 53, and when from equation 61 the following relation is realized

$$\frac{E_{c0}}{1+H_c/R_0}>e_{ac}' \tag{67}$$

the current i_1 again flows in the gate circuit, and the mode changes to mode 0. The phase angle, θ_2, of transition from mode III to mode 0 is found from the following relation

$$\frac{E_{c0}}{1+R_c/R_0}=E'\sin\theta_2 \tag{68}$$

These relations concerning the sequence of the modes are represented by Fig. 11.

Comparison with the Basic Circuit

In comparing the modes of the basic circuit shown in Fig. 5 with those in Figs. 9 and 10, the difference due to commutation is found only in mode I and 0. Therefore, the amount of reset flux during the interval from 0 to θ_1, designated as $\Delta^2\Phi(\theta_1)$, can be calculated from equations 57 and 60 as follows:

$$N\omega\Delta^2\Phi(\theta_1)=\frac{E_{c0}[\theta_1\sin\theta_1-(1-\cos\theta_1)]}{(1+R_c/R_0)\sin\theta_1}\times \frac{R_c/R_{LG}}{1+R_c/R_0+R_c/R_{LG}} \tag{69}$$

and the amount during the interval from θ_2 to π is

$$N\omega\Delta^2\Phi(\Delta\theta_2)= \frac{E_{c0}[\Delta\theta_2\sin\Delta\theta_2-(1-\cos\Delta\theta_2)]}{(1+R_c/R_0)\sin\Delta\theta_2}\times \frac{R_c/R_{LG}'}{1+R_c/R_0+R_c/R_{LG}'} \tag{70}$$

where

$$\Delta\theta_2=\pi-\theta_2 \tag{71}$$

Using the expression

$$\sigma(\theta)=\frac{\theta\sin\theta-(1-\cos\theta)}{\pi\sin\theta} \tag{72}$$

$$\rho=\frac{R_c/R_{LG}}{1+R_c/R_0+R_c/R_{LG}} \tag{73}$$

$$\rho'=\frac{R_c/R_{LG}'}{1+R_c/R_0+R_c/R_{LG}'} \tag{74}$$

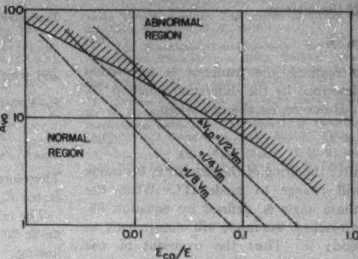


Fig. 14. Critical condition for normal modes

and putting

$$\Delta^2\Phi=\Delta^2\Phi(\theta_1)+\Delta^2\Phi(\Delta\theta_2) \tag{75}$$

$$\Delta^2\Phi=\frac{E_{c0}}{2fN(1+R_c/R_0)}[\sigma(\theta_1)\rho+\sigma(\Delta\theta_2)\rho'] \tag{76}$$

the fundamental difference equation of the full-wave magnetic amplifier is, from equations 37, 5, 56, and 76, as follows:

$$\Delta\Phi(n)=\frac{\Delta\Phi(n-1)}{1+R_c/R_0}+\frac{1}{N(1+R_c/R_0)}\times \frac{E_{c0}}{2f}-\Delta^2\Phi \tag{77}$$

that is,

$$\Delta\Phi(n)=\frac{\Delta\Phi(n-1)}{1+R_c/R_0}+\frac{E_{c0}}{2fN(1+R_c/R_0)}\times \{1-\sigma(\theta_1)\rho-\sigma(\Delta\theta_2)\rho'\} \tag{78}$$

In equation 78, when R_c is comparatively small, ρ and ρ' are negligible, as seen from equations 73 and 74, and equation 78 reduced to that of the basic circuit, equation 37.

In practice

$$R_0>>R_{LG} \tag{79}$$

and

$$R_0>>R_{LG}' \tag{80}$$

Therefore, when the conditions

$$R_c>>R_{LG} \tag{81}$$

and

$$R_c>>R_{LG}' \tag{82}$$

are satisfied, from equations 73 and 74

$$\rho=\rho'=1 \tag{83}$$

From equation 78 it may be said that the time constant of the full-wave magnetic amplifier is not affected by the commutation. But for the case of high control resistance, the steady-state value of the reset flux is decreased, compared to that of the basic magnetic amplifier of Fig. 5, as seen from equations 78 and 15(A).

The functions $\sigma(\theta_1)$ and $\sigma(\Delta\theta_2)$ in equation 78 are the important factors which represent the influence of commutation. From equation 64, $\sin\theta_1$ is related directly to the output voltage of the basic circuit, i.e.,

$$A_{v0}E_{c0}/E=\sin\theta_1 \tag{64}$$

Therefore, $\sigma(\theta)$ versus $\sin\theta$ is represented in Fig. 12. As an example, Fig. 13 shows the effect of commutation on the control characteristics of the full-wave amplifier influenced by $\sigma(\theta)$, i.e., under the assumption of

$$\left.\begin{array}{l}\rho=1\\ \sigma(\Delta\theta_2)\rho'=0\end{array}\right\} \tag{84}$$

However, this is not complete enough to explain the actual change of the control characteristics due to the change of control circuit resistance.[10] Therefore, the change of the more complex dynamic behavior of the core should be considered.[9,10]

Initial Portion of the Switch-in Transient

Assume, at first, that the control voltage is zero and both cores are saturated. Then apply the control step function at the beginning of the gating half-cycle of core 1. Thus, core 1 will continue to be saturated until the end of the first half-cycle, and the flux of the core 2 will be reset, in sequence of modes 0, III, and 0 by the following amount:

$$2fN\Delta\Phi(1)=\frac{E_{c0}}{1+R_c/R_0}[1-2\rho'\sigma(\theta_0)] \tag{85}$$

where

$$\sin\theta_0=\frac{E_{c0}}{E'(1+R_c/R_0)} \tag{86}$$

On the second half-cycle, the flux of core 2 is reset, and first mode I is realized. The firing angle, α_2, of the reactor 2 is given by

$$1-\cos\alpha_2=\frac{N\omega}{E}\Delta\Phi(1) \tag{87}$$

If the relation

$$\alpha_2 > \theta_1 \tag{88}$$

is realized, the transient is completely described by the difference equation 78.

However, for the case when

$$\alpha_2 < \theta_1 \tag{89}$$

after the firing of the reactor 2, the mode will change to mode III. When the phase angle θ_2, defined by equation 68, is reached, the mode will change back to mode 0. Thus the transient of the modes is different from the previous case, and this sequence can be considered as the abnormal modes of transition.

To seek the fundamental difference equations corresponding to the abnormal modes, the amount of reset flux before the firing angle α_n is calculated, from equation 57, as follows:

$$\Delta\Phi'(n) = \frac{E_{c0}\alpha_n}{N\omega(1+R_c/R_0+R_c/R_{LG})} + \frac{1+R_c/R_{LG}}{1+R_c/R_0+R_c/R_{LG}}\Delta\Phi(n-1) \tag{90}$$

From equations 61 to 70, the amount of reset flux from the instant of the firing of the core to the end of the resetting half-cycle, i.e., during the modes III and 0, is

$$\Delta\Phi''(n) = \frac{E_{c0}(\pi-\alpha_n)}{N\omega(1+R_c/R_0)} - \Delta^2\Phi(\Delta\theta_2) = \frac{E_{c0}}{2fN(1+R_c/R_0)}\left[1-\frac{\alpha_n}{\pi}-\rho'\sigma(\Delta\theta_2)\right] \tag{91}$$

From equations 90, 91, and 73, the resultant difference equation is

$$\Delta\Phi(n) = \frac{1+R_c/R_{LG}}{1+R_c/R_0+R_c/R_{LG}}\Delta\Phi(n-1) + \frac{E_{c0}}{2fN(1+R_c/R_0)}\left\{1-\frac{\alpha_n}{\pi}-\rho'\sigma(\Delta\theta_2)\right\} + \frac{E_{c0}\alpha_n}{2fN(1+R_c/R_0)\pi}(1-\rho)$$

$$= \frac{1+R_c/R_{LG}}{1+R_c/R_0+R_c/R_{LG}}\Delta\Phi(n-1) + \frac{E_{c0}}{2fN(1+R_c/R_0)}\left\{1-\frac{\alpha_n}{\pi}\rho-\rho'\sigma(\Delta\theta_2)\right\} \tag{92}$$

This is more complicated than equation 12 because α_n is a function of $\Delta\Phi(n-1)$, and cannot be explained by the simple time constant. If the relation

$$\frac{\alpha_n}{\pi}\rho << 1 \tag{93}$$

is satisfied, the time constant for the abnormal modes will be

$$T_r = \frac{1}{2f}\frac{1}{\ln\left(1+\dfrac{R_c/R_0}{1+R_c/R_{LG}}\right)} \tag{94}$$

and when

$$R_c/R_0 << 1 \tag{95}$$

$$T_r = \frac{1}{2f}\frac{R_0}{R_c}\left(1+\frac{R_c}{R_{LG}}\right) \tag{96}$$

Therefore the response is retarded by the factor R_c/R_{LG}.

The critical condition for the abnormal mode to appear is discussed in the following. The critical condition is defined by the relation

$$\alpha_2 = \theta_1 \tag{97}$$

From equations 85, 87, and 97 using equations 64 and 44

$$1-\cos\theta_1 = \left(\frac{\pi}{1+A_{r0}}\right)(\sin\theta_1\{1-2\rho'\sigma(\theta_0))\} \tag{98}$$

Neglecting the factor $2\rho'\sigma(\theta_0)$

$$A_{r0} = \frac{\pi\sin\theta_1}{1-\cos\theta_1} - 1 \tag{99}$$

is obtained. Using equations 64 and 99, the relation between E_{c0}/E and A_{r0} for the critical condition is obtained as shown in Fig. 14. In this figure, when the condition of the amplifier corresponds to the point in the abnormal region, the abnormal mode will appear at the beginning of the switch-in transient. Yet the appearance of this mode is far shorter than the normal modes governed by equation 78. Therefore, the time constant of the full-wave magnetic amplifier can usually be represented by a single time constant.

Conclusions

The general nature of the first-order linear difference equation is first discussed because of its importance in the analysis of the magnetic amplifier circuits. Thereafter the difference equation for the basic full-wave magnetic amplifier circuit is derived using the mathematical representation of the core characteristics, which takes into account the change in width of the dynamic hysteresis loop.

The results of solving the difference equation, explains the transient and steady-state behavior of the full-wave magnetic amplifier circuit fairly well.

The effect that commutation in the gate circuit has on the operation of the amplifier is also discussed. For the normal case, it is made clear that only the steady-state output is changed by a small amount without any change of the time constant. The possibility of the existence of the abnormal modes where the system is represented by the nonlinear difference equation is pointed out.

Appendix. Duration of Mode I

The circuital equation for mode I is

$$\begin{Vmatrix} R_L+R_G & R_L & 0 & 1 & 0 \\ R_L & R_L+R_G & 0 & 0 & 1 \\ 0 & 0 & R_c & -1 & -1 \\ 1 & 0 & -1 & -k/N^2 & 0 \\ 0 & 1 & -1 & 0 & -k/N^2 \end{Vmatrix} \begin{Vmatrix} i_1 \\ i_2 \\ i_c \\ Nd\Phi_1/dt \\ Nd\Phi_2/dt \end{Vmatrix} = \begin{Vmatrix} -e_{ac} \\ e_{ac} \\ E_c \\ --i_0 \\ i_0 \end{Vmatrix} \tag{100}$$

Calculating the current i_1 from equation 100, and seeking the value of phase angle θ_1 when i_1 goes to zero,

$$\sin\theta_1 = \left(\frac{E_{c0}A_{r0}}{E}\right)\times\left(\frac{1-\dfrac{i_0(R_G+2R_L)}{E_{c0}}(2+A_{r0}^{-1})}{1+\dfrac{R_G+2R_L}{R_c}(2+A_{r0}^{-1})}\right) < A_{r0}\frac{E_{c0}}{E} \tag{101}$$

under the assumption that

$$R_G/R_0 << 1 \tag{102}$$

This value of θ_1 is usually smaller than the value given by equation 64. This means that the duration of mode I is usually shorter than that assumed by mode I′, therefore the effect of the commutation is smaller for the amount of reset flux. Consequently, the case discussed in the paper can be considered to be the upper limiting case, but in spite of this the general tendency of the change of the modes is explained sufficiently.

References

1. THEORY OF MAGNETIC AMPLIFIERS WITH SQUARE-LOOP CORE MATERIALS, H. F. Storm. *AIEE Transactions*, vol. 72, pt I, Nov. 1953, pp. 629–40.

2. RESPONSE TIME OF MAGNETIC AMPLIFIERS, E. L. Harder, W. F. Horton. *Ibid.*, vol. 69, pt. II, 1950, pp. 1130–41.

3. ANALYSIS OF MAGNETIC AMPLIFIERS BY THE USE OF DIFFERENCE EQUATIONS, P. R. Johannessen. *Ibid.*, vol. 71, pt. I, 1954 (Jan. 1955 section), pp. 700–11.

4. ON THE MECHANICS OF MAGNETIC AMPLIFIER OPERATION, Robert A. Ramey. *Ibid.*, vol. 70, pt. II, 1951, pp. 1214–23.

5. THE SERIES MAGNETIC AMPLIFIER, PART I–THE FOUR MODES OF OPERATION; PART II–THE MECHANISMS OF GAIN AND TIME CONSTANT, R. C. Barker. *Ibid.*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 819–31.

6. THE SIMPLE REACTOR CIRCUIT; ITS OPERATION

AND MODE TRANSITION, J. F. Ringelman, A. L. Fenaroli. *Ibid.*, pp. 660–69.

7. EFFECTS OF THE CIRCUIT PARAMETERS UPON THE TRANSIENT CHARACTERISTICS OF THE MAGNETIC AMPLIFIER, T. Anayama. *Technical Paper No. 273* (in Japanese). *Proceedings*, Institute of Electrical Engineers, Tokyo, Japan, May 1955.

8. ON THE MECHANICS OF SELF-SATURATING MAGNETIC AMPLIFIER OPERATION, T. Kusuda. *Journal*, Institute of Polytechnics, Osaka City University, Osaka, Japan, Series E, vol. 3, pp. 103–17.

9. THE INFLUENCE OF MAGNETIC AMPLIFIER CIRCUITRY UPON THE OPERATING HYSTERESIS LOOPS, Harold W. Lord. *AIEE Transactions*, vol. 72, pt. I, 1953 (Jan. 1954 section), pp. 721–28.

10. ON THE FLUX CONTROL OF THE SATURATING REACTOR FOR MAGNETIC AMPLIFIERS, T. Kikuchi, K. Murakami. *Technology Reports*, Tohoku University, Sendai, Japan, vol. 20, 1956, pp. 165–95.

11. ON THE INPUT POWER OF THE HALF-WAVE MAGNETIC AMPLIFIER CIRCUIT, T. Kikuchi, K. Murakami. *AIEE Transactions*, vol. 76, pt. I, Mar. 1957, pp. 10–14.

12. SYNTHESIS OF HALF-WAVE MAGNETIC AMPLIFIER CIRCUITS WITH DECREASED INPUT POWER, T. Kikuchi. *Proceedings*, Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., 1956.

13. THE RESIDUAL TIME-CONSTANT OF SELF-SATURATING TRANSACTORS, Ulrik Krabbe. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 103, part C, pp. 71–80, 1956.

14. DESIGN FACTORS FOR MAGNETIC AMPLIFIERS, K. W. Gardiner. *Technical Report 7002-9*, Servomechanisms Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., Aug. 31, 1954.

15. MAGNETIC AMPLIFIERS (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, p. 310.

The Clamp-Type A-C Microammeter

G. FRANKLIN MONTGOMERY
NONMEMBER AIEE

CARROLL STANSBURY
FELLOW AIEE

FOR MANY YEARS, the range of current-indicating instruments at power frequencies has been extended by the use of current transformers. A particularly convenient method is that in which the transformer consists of a secondary winding on a split toroidal core that can be clamped around the conductor whose current is to be measured. This technique suggests a similar device for measuring small currents at audio and radio frequencies: an instrument consisting of a small clamp-type magnetic probe and suitable means for amplifying the output of the probe to operate an indicating meter or an oscilloscope. Such an instrument is complementary to the a-c voltmeter in the audio- and radio-frequency range and has obvious utility in experimental work; this paper will demonstrate its realization.

Theory

Suppose that the current to be measured is sinusoidal, has rms amplitude I amperes and frequency $f=\omega/2\pi$ cps (cycles per second), and is carried by a conductor that passes through the window of a continuous magnetic core. An N-turn winding on the core has self-inductance L henries and a-c resistance R_s ohms. A diagram of the arrangement is shown in Fig. 1, to the left of terminals c and d.

The quantity of interest is the voltage induced in the N-turn winding by the current I. This voltage can be obtained by calculating the magnetic flux in the core from the Biot-Savart law, integrating the flux over the core cross section, and differentiating the result to find the induced voltage. A simpler approach is to note that the conductor and winding form the primary and secondary windings of a transformer of turns ratio N. Then, if terminals a and b could be short-circuited, the short-circuit current, neglecting the secondary leakage reactance, would be

$$I/N=E/\omega L \qquad (1)$$

where E, in volts, is the rms voltage induced by current I. The self-inductance, however, is

$$L=N^2\mathcal{P} \qquad (2)$$

where $\mathcal{P}$ is the permeance of the core in webers per ampere-turn, so that

$$E=\omega LI/N=\omega N\mathcal{P}I \qquad (3)$$

The quantity $N\mathcal{P}$ is the mutual inductance of the transformer.

THE SIGNAL

The object in a practical instrument is to use the secondary output to actuate an indicating meter; if the current to be measured is small, the secondary output will require amplification before it can do so. It is also evident that if the secondary load impedance is large compared with its self-reactance, the load current will be proportional to the induced voltage, and, therefore, to frequency. The most useful instrument, however, is one in which the indication is independent of frequency. Both the required amplification and frequency independence can be achieved in the following way.

In Fig. 1, terminals c and d are the input terminals of an amplifier whose input resistance is R_a. E_n is a zero-impedance voltage generator that represents the internally developed amplifier noise. The amplifier input signal current is

$$I_s=\frac{E}{\sqrt{R^2+(\omega L)^2}} \qquad (4)$$

where $R=R_s+R_a$. By substitution from equation 1,

$$I_s=\frac{I/N}{\sqrt{1+(f_0/f)^2}} \qquad (5)$$

where $f_0=R/2\pi L$. Now, let the current gain of the amplifier be

$$A=A_0\sqrt{1+(f_0/f)^2} \qquad (6)$$

in the range $f_1 \leqslant f \leqslant f_2$ and zero elsewhere. The amplifier signal output current will then be

$$I_{s0}=AI_s=A_0I/N \qquad (7)$$

which is independent of frequency over the specified range.

THE NOISE

The constitution of the internally developed noise can be represented by

$$E_n^2=\int_{f_1}^{f_2}e_n^2df \qquad (8$$

It is certain that for any practical amplifier, e_n^2 will be a function of frequency. For those amplifiers that are close to ideal, however, e_n^2 will be nearly constant; it will be assumed that e_n^2 is constant and defines an R_n such that

$$e_n^2=4kTR_n \qquad \text{(volt}^2\text{-seconds)} \qquad (9)$$

where k is Boltzmann's constant (1.38×10^{-23} joules per degree Kelvin), and T is the temperature in degrees Kelvin. It is likewise certain that R_s will vary with frequency, but to simplify matters it will be assumed that R_s can be represented adequately by an average value. With these assumptions, the differential input noise current is given by the following equation

---

Paper 57-114, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 8, 1956; made available for printing December 6, 1956.

G. FRANKLIN MONTGOMERY and CARROLL STANSBURY are with the National Bureau of Standards, Washington, D. C.

This work was conducted under a program of basic instrumentation research and development jointly sponsored by the Office of Naval Research, the Office of Scientific Research of the Air Research and Development Command, and the Atomic Energy Commission.

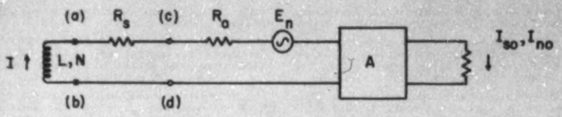


Fig. 1. Input circuit of ideal instrument

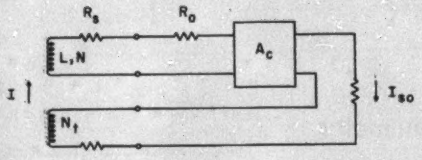


Fig. 2. Probe input circuit with feedback

$$dI_n^2 = \frac{4kT(R_s+R_n)df}{R^2+(\omega L)^2} \tag{10}$$

which, from the definition of f_c, becomes

$$dI_n^2 = \frac{kT(R_s+R_n)df}{(\pi L)^2[1+(f_c/f)^2]f^2} \tag{11}$$

Using equation 6, the square of the amplifier noise output current will be

$$I_{no}^2 = \int_{f=f_1}^{f_2} A^2 dI_n^2 = \frac{A_0^2 kT(R_s+R_n)}{(\pi N^2\mathcal{P})^2} \times (1/f_1 - 1/f_2) \tag{12}$$

THE SIGNAL-TO-NOISE RATIO

Now, from equations 7 and 12, the output signal-to-noise power ratio is

$$S = I_{so}^2/I_{no}^2 = \frac{(\pi N\mathcal{P}I)^2 f_1 f_2}{kT(R_s+R_n)(f_2-f_1)} \tag{13}$$

But for most practical instruments, $f_2 >> f_1$, so that

$$S = \frac{f_1(\pi N\mathcal{P}I)^2}{kT(R_s+R_n)} \tag{14}$$

approximately. This is the basic design equation for the instrument; note that it is independent of f_2, the upper frequency limit of the band.

Circuit Design

It has been supposed that the amplifier gain varies with frequency according to equation 6. If this equalization were provided by a network within the amplifier, both the calibration of the instrument and its frequency response would depend upon the permeance of the core; for a split core, some means would be necessary to ensure that the core-gap permeance was the same for each measurement.

A better equalization method is to use negative current feedback from the output of the amplifier to the secondary winding itself. The feedback current will develop across the secondary reactance a counter electromotive force that opposes the induced secondary voltage and is proportional to the core permeance. This action is equivalent to loading the secondary with a low resistance; if the resistance is sufficiently small, it approximates a short circuit, and equation 1 becomes applicable in fact. Since both the secondary induced voltage and the secondary reactance are proportional to permeance, the short-circuit current is independent of permeance and of physical change in the gap.

The path of the feedback current must not include the secondary resistance R_s, however, for the voltage drop across this resistance is in series with the input to the amplifier. To avoid this difficulty, feedback may be applied to a tertiary winding on the core. If unity coupling is assumed between the secondary and tertiary windings, then, in Fig. 2,

$$\mathbf{I}_{so} = A_c\mathbf{I}_s = \frac{A_c[j\omega L\mathbf{I}/N - j\omega L\mathbf{I}_{so}(N_t/N)]}{R+j\omega L} = \frac{A_c(\mathbf{I}-N_t\mathbf{I}_{so})}{N(1-jf_c/f)} \tag{15}$$

where A_c = (constant) is the amplifier current gain in the frequency range $f_1 \leqslant f \leqslant f_2$, and N_t is the number of turns on the tertiary winding. Then,

$$\mathbf{I}_{so}(1-jf_c/f+A_cN_t/N) = A_c\mathbf{I}/N \tag{16}$$

and

$$I_{so} = \frac{I/N}{\sqrt{(1/A_c+N_t/N)^2+(f_c/A_cf)^2}} \tag{17}$$

If $A_cN_t/N >> 1$, then

$$I_{so} = \frac{I/N}{\sqrt{(N_t/N)^2+(f_c/A_cf)^2}} \tag{18}$$

approximately. It is evident from comparison of equations 5 and 18 that feedback extends the range of constant response downward in frequency; the new cutoff frequency is

$$f_c' = Nf_c/N_tA_c \tag{19}$$

It can be shown that for a given f_c' and a given core, the necessary A_c will be least when $N_t = N$, approximately. It therefore will be assumed that

$$f_c' = f_c/A_c \tag{20}$$

In general, A_c will be limited by the maximum gain that can be obtained in a 2-stage amplifier; feedback over a larger number of stages inevitably causes instability at one of the extremes of the frequency range unless complicated networks are included to control the loop phase shift. The necessary A_c is proportional to f_c, which increases with $R = R_s + R_a$. It is therefore advantageous to arrange the amplifier so that its input resistance, R_a, is as small as possible. One method of doing so is to transformer-couple the secondary winding to the amplifier input. This expedient increases the sensitivity, for the N of equation 14 is then multiplied by the transformer turns ratio. R_s is also multiplied by the square of the turns ratio, however, so that the improvement in sensitivity by this means has an upper limit. An important characteristic of the transformer in this application is its winding resistance; if the winding resistances greatly exceed their corresponding load impedances at the low-frequency end of the range, the improvement sought by transformer coupling will be cancelled. In addition, the parasitic

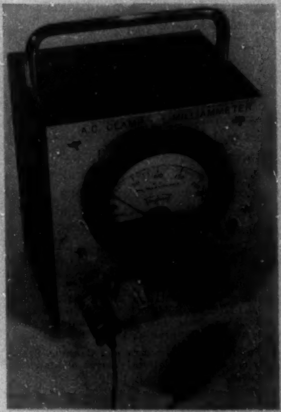


Fig. 3. Experimental 200-microampere instrument

reactances of the transformer may cause instability when it is included in the amplifier feedback loop. A large improvement, therefore, can probably be obtained by transformer coupling only when the frequency range of the instrument is to be relatively narrow.

An Example

An experimental instrument, shown in Fig. 3, has been constructed not as a demonstration of ultimate performance in either sensitivity or frequency range but rather as a technical exercise to test the conclusions of theory. The clamp-type probe is shown in the same illustration.

THE PROBE

Equation 14 reveals the importance of both the number of turns of wire on the core and its permeance. For large sensitivity, the core should be made of material having large relative permeability and should have a large ratio of cross-section area to length of flux path. To accommodate a large N, the window area should be as large as possible. To be generally useful in confined spaces, the over-all dimensions should be small. These requirements are evidently contradictory and require the usual vexing compromise.

For the experimental instrument, a square core was chosen of 0.014-inch silicon-steel laminations formed in matching C's and I's. Assembled, the core dimensions are

Thickness (stack height) = 1/8 inch
Width = height = 5/16 inch
Window width = window height = 1/8 inch

The legs of the core C project upward from the lower jaw of the Bakelite clamp, in which the C and its windings have been cemented; the core I is contained in the upper jaw. A pin passes through the center of the I, allowing rotary movement in the plane of the jaws, so that as the jaws close the I will seat itself properly on the projecting legs of the C. The matching faces of these parts are ground.

The core C contains two windings, each having 250 turns of number 44 enameled wire. The four coil leads connect to a shielded 4-wire cable and plug. The electrical characteristics for each winding are

$L = 1.4\times10^{-2}$ henries
$Q = 1.9$ (at 1,000 cps)
D-c resistance = 30 ohms
Resonant frequency (with cable) = 1.2×10^{5} cps

From equation 2, the core permeance is

$$\mathcal{P} = L/N^2 = 2.2\times10^{-7} \text{ webers per ampere-turn} \qquad (21)$$

The permeance of the core could have been calculated directly from its physical dimensions, knowing the relative permeability μ and making an appropriate estimate of the flux path length. Working backward from the given permeance, with physical dimensions in meters and $\mu_0 = 4\times10^{-7}$ henries per meter,

$$\mu = \frac{\mathcal{P}\times\text{path length}}{\mu_0\times\text{cross-section area}} = 516 \qquad (22)$$

which is reasonable for the material used, including the gap.

SENSITIVITY

A lower cutoff frequency $f_1 = 20$ cps and a signal-to-noise power ratio $S = 10^4$ are now selected, and the minimum current I corresponding to these specifications calculated. Values must also be chosen for R_s, R_n, and T; assume that $R_s + R_n = 10^3$ ohms, and $T = 300$ degrees Kelvin. Then, from equation 14,

$$I = \frac{1}{\pi N\mathcal{P}}\sqrt{\frac{kT(R_s+R_n)S}{f_1}} = 2.6\times10^{-4} \text{ amperes} \qquad (23)$$

This value suggests an instrument with a full-scale sensitivity of 200 microamperes having a noise level slightly greater than one per cent of full scale.

THE CIRCUIT

In any design of this type, it is helpful to work backward from the output end of the amplifier toward the input. An indicator having a basic 200-microampere d-c movement is a convenient compromise between cost, sensitivity, and ruggedness. To drive this instrument to full scale through a full-wave rectifying network will require an rms current of

$$I_s = 2\times10^{-4}/0.9 = 2.22\times10^{-4} \text{ amperes} \qquad (24)$$

where the 0.9 denominator is the ratio of average to rms value of a sine wave. The rectifier used will inevitably be nonlinear for small input currents, but the nonlinearity will be reduced by driving the rectifier with an amplifier whose gain is sacrificed to provide negative feedback through the rectifier. In Fig. 4, transistor $V7$ is connected in such an arrangement. The full-scale current input to $V7$ is approximately that given by equation 24.

The preamplifier is defined as that part of the circuit which provides the equalizing feedback current to the tertiary winding of the probe; in Fig. 4 transistors $V1$ and $V2$ comprise the preamplifier. From equation 18, the preamplifier output current in the range $f >> f_0$ is, if $N = N_t$ as it is in this case,

$$I_m \cong I/N = 8\times10^{-7} \text{ amperes} \qquad (25)$$

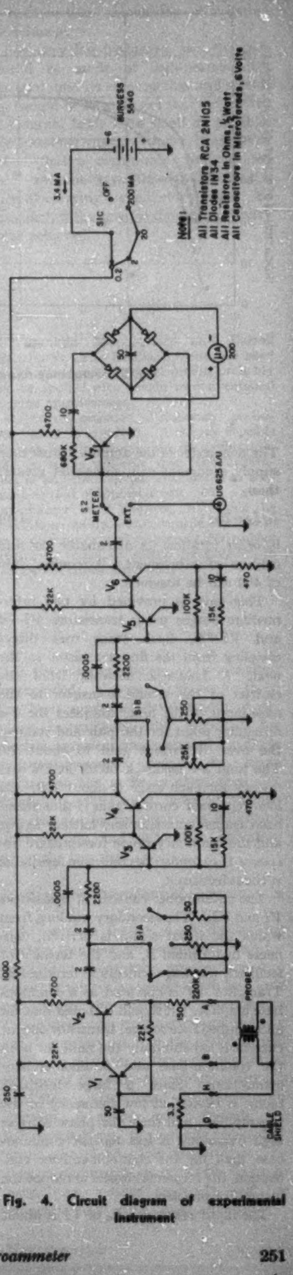


Fig. 4. Circuit diagram of experimental instrument

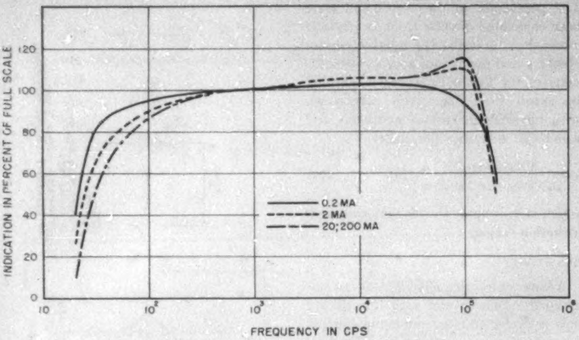


Fig. 5. Frequency response of experimental instrument

The remainder of the amplifier must then supply a current gain somewhat greater than

$$I_0/I_{60}=280 \qquad (26)$$

in order to allow for attenuator loss and calibration adjustment. A required gain of 400 will be assumed.

This gain is provided by two intermediate stages using transistors *V3-V4* and *V5-V6*. Each stage uses direct coupling from the first transistor to the next. A feedback network, from the emitter of the second transistor to the base input of the first, stabilizes the d-c operating point for the pair and reduces the over-all current gain to about 20. The total a-c feedback factor at low frequency for each stage is about 100; the low-frequency current gain is thus stabilized against transistor and battery aging, and the frequency range for constant response is extended beyond that available without feedback.

The preamplifier consists of transistors *V1* and *V2*. The secondary winding, from which the input signal is derived, connects to terminal *B*, and the tertiary or feedback winding connects to terminal *A*. Transistor *V1* is operated as a common-base amplifier; it should be noted that for a low-impedance source, transistor output current is substantially the same for both the common-base and common-emitter connections. Since a large feedback factor is needed at low frequency in the preamplifier, and since the phase shift at high frequency is less for the common-base than for the common-emitter connection, the former is chosen to reduce the instability problem.

The input resistance R_a of *V1* is about 100 ohms. Then, in the definitions of equations 4 and 5, $R=130$ ohms, and

$$f_c=R/2\pi L=1.5\times 10^3 \text{ cps} \qquad (27)$$

But a cutoff frequency of 20 cps is desired. Therefore, from equation 20,

$$A_c=f_c/f_c'=75 \qquad (28)$$

This gain is greater than can be obtained ordinarily with the transistors used by a factor of about 3. To make up the deficiency, positive current feedback is included within the negative loop by means of adding the 3.3-ohm resistor at terminal *H*.

Results

Response curves of the instrument are given in Fig. 5. The characteristic of particular interest is that for the 0.2-milliampere range; the additional low-frequency attenuation shown for the higher current ranges is not fundamental but is due to insufficient capacitance in the 2-microfarad interstage coupling capacitors. The variation between ranges at high frequency is fundamental and illustrates the difficulty of designing simple attenuators that will not affect the high-frequency response of transistors used in feedback circuits. What is needed here is some form of constant-impedance attenuator rather than the simple current shunts that are used.

Linearity of response is within the 2-per-cent specification of the indicating instrument itself, showing that the degree of feedback used in *V7* is adequate. Less feedback might be permissible if diodes were used having greater forward conductance than the *1N34*.

The noise level of the instrument is about 20 microamperes. This result is surprising, particularly as it has been determined that the noise level generated by the input transistor itself does not exceed that assumed in the calculations. The input noise spectrum has a maximum below 50 cps. One possible explanation is the production of Barkhausen noise within the core material by low-frequency components of the emitter currents from *V1* and *V2;* this conjecture has not been proved.

It was soon discovered that fields within the laboratory at the 60-cps power frequency and its harmonics up to at least the fourth are a source of interference. At a typically "quiet" location of the probe on the laboratory bench, the equivalent current generated by hum is about 30 microamperes. Such fields are a real problem in attempting to make measurements in the vicinity of power transformers, etc. Two mumetal shield caps, not shown in Fig. 3, made of $^1/_{32}$-inch material and properly annealed, were fitted over the clamp jaws in a manner that does not interfere with normal use of the clamp. They are effective in reducing the hum pickup by a factor of about 4.

Insertion Impedance

Whenever current is measured, an impedance is effectively inserted in series with the circuit carrying the measured current. The impedance introduced by the clamp-type instrument will be calculated in the following.

The combination of probe and preamplifier is equivalent to a lossless probe that delivers I/N amperes to a fictitious resistive load. The resistance of the load is

$$R_f=2\pi f_c'L \qquad (29)$$

Viewing the probe as a transformer, the impedance at the input or primary "terminals" is formed by an inductance L_p in parallel with a resistance R_p, where, from equation 2,

$$L_p=L/N^2=\mathcal{P} \qquad (30)$$

and

$$R_p=R_f/N^2=2\pi f_c'\mathcal{P} \qquad (31)$$

It is evident that over most of the useful frequency range, the insertion impedance will be almost entirely resistive.

For the instrument used in illustration,

$$L_p=2.2\times 10^{-7} \text{ henries} \qquad (32)$$

and

$$R_p=2.8\times 10^{-5} \text{ ohms} \qquad (33)$$

Conclusions

It has been shown that an instrument using a clamp-type magnetic probe can be used to measure small alternating currents in the audio- and low radio-frequency ranges and that such an instrument introduces a low impedance in the measured circuit. In general, the experimental work has confirmed the initial theory; unforeseen results are the higher-than-predicted noise level and the severity of power-frequency hum interference. It is logical, therefore, that successful instruments of this type may be designed for other frequency ranges and other sensitivities.

In the experimental instrument, the high-frequency limit is imposed mainly by limitations of the transistors used, although for the particular probe that has been discussed, the maximum operating frequency is probably not more than three or four times its resonant frequency of 120 kc. Instruments designed particularly for high-frequency use would well make use of more appropriate core material and a construction minimizing distributed capacitance.

For very low-frequency use, the core should be made of high-permeability material, and its construction and shielding should be so arranged that sensitivity to external fields is as small as possible. For power-frequency work, it may be found that a continuous, well-shielded core, without gap, through which the current-carrying conductor must be threaded, will allow maximum sensitivity to be achieved.

◆

Discussion

Francis L. Hermach (National Bureau of Standards, Washington, D. C.): The authors have modestly refrained from emphasizing their achievement in designing and constructing an instrument using a current transformer that operates on only 200 microampere-turns full scale. Generally, a minimum of 100 to 1,000 ampere-turns are required for good performance in ordinary 60-cycle transformers of the precision class, about 1,000,000 times their value. Perhaps 10 ampere-turns are required for a hook-on ammeter using a current transformer and a rectifier instrument of the accuracy class of their instrument, and only a very narrow frequency range is usually obtainable. This points out the rather dramatic advantages which result from this development in feedback amplifiers utilizing modern components.

It would be interesting to know if similar techniques would make possible improved accuracy, wider frequency range, or smaller size in current transformers of the more usual current ranges. A wider frequency range would be especially interesting, as the authors quite confidently expect operation even at frequencies considerably above the resonant frequency of the transformer.

G. Franklin Montgomery and Carroll Stansbury: The feedback technique used in the microammeter development probably can be applied with benefit to conventional current transformers.

As to the question of accuracy, no improvement is expected. Accuracy depends ultimately on maintenance of calibration, and it seems unlikely that the calibration of an active device will remain as constant as that of a passive one. For a given transformer size, however, it should be possible to obtain a considerable increase in frequency range, and conversely, for a given frequency range, it should be possible to reduce the transformer size.

A New Voltage Time Recorder for Transient Measurements

C. M. FOUST
FELLOW AIEE

B. V. BHIMANI
ASSOCIATE MEMBER AIEE

ENGINEERING measurements of electric circuit quantities such as voltage, current, power, energy, and wave shapes have, for the steady-state conditions including both non varying quantities and recurring waves, been made with general satisfaction, through steadily improving designs of voltmeters, ammeters, wattmeters, and oscillographs. These are measurements of the quantities directly pertinent to the purposes for which the circuitry was designed. All circuitry, however, is also unintentionally subjected to electric transients when steady-state levels are changed and when abnormal electrifications or connections occur. These are mainly switching, lightning, and insulation failure transients and they involve voltage and current variations where fractions of a microsecond sometimes become important because of the high magnitudes of the voltages and currents involved.

For electric transient measurements, particularly in the range of more rapid changes, the steady-state instruments and meters are inappropriate because, in all cases except some oscillographs and oscilloscopes, they have response times longer than the electrical changes to be measured or recorded. Necessarily therefore there has been a parallel development of devices, instruments, and recorders for electric transient measurements. These transient-measuring devices can incorporate far less accuracy than those for steady state but must, however, include the features of instantaneous indicating or recording and automatic operation when the transient occurs.

Devices for Transient Measurements

A complete listing of all devices for electric transient measurements both for specialized laboratory applications and general engineering usage would be a long one. Since the instrument described in this paper is for engineering use both in the laboratory and in the field, it will suffice to relate it to the most important devices of the same classification already available and developed over the last few decades. A listing must include the sphere spark gap, the cathode ray-oscillograph, the powder link flashover indicators, the transient analyzer, the Lichtenberg figure voltage recorders, and the magnetic link surge crest ammeter.

Paper 57-214, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted June 20, 1956; made available for printing December 17, 1956.

C. M. Foust and B. V. Bhimani are with the General Electric Company, Schenectady, N. Y.

The assistance of N. Rohats in the design of the recorder is gratefully acknowledged. P. N. Bosworth contributed in helpful suggestions and a review of the paper, and Stanley Locke and John Adams in making experimental calibrations.

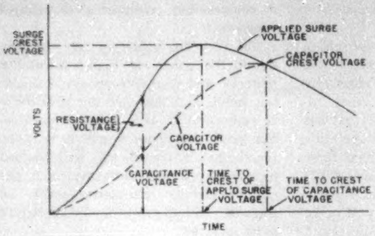


Fig. 1. Resistance-capacitance voltage division

Sphere Spark Gap

The calibration of the sphere spark gap, in 1913, provided electrical engineers with the first device for quantitative measurements of transient voltages.[1-3] General acceptance of the sphere gap resulted, when calibrations from several laboratories were found to be in good agreement. Laboratory use of the sphere gap spread rapidly as applications multiplied and its conveniences, such as low capacitance and physical simplicity, were recognized. Its principal limitation has been that, to measure a transient voltage, several gap spacings must be exposed to the same crest voltages for sparkover observations. This necessary repeating of the transient to be measured has disqualified the sphere spark gap for single transient voltage measurement and rendered it inappropriate for recording unpredictable transient voltages. Thus the place for the sphere gap, as a voltage measurement device, is almost exclusively the laboratory where the same transient or wave shape can be repeated until its crest voltage is determined through adjusted gap spacings.

Cathode-Ray Oscillograph

This nearly perfect and widely adaptable instrument was successfuly applied to the measurement of short-time transients more than three decades ago as described in the technical literature in 1922.[4] The first high-speed cold-cathode single-trace oscillograph in this country was operated by one of the authors in 1924. The appropriateness of the cathode-ray beam for the recording of extreme rates of change of voltage or current was quickly recognized and many applications in the laboratory and in the field followed rapidly. The early cold cathode-ray oscillographs required continuous evacuation to preserve the proper cathode tube pressures for beam focusing. In 1936, this problem was solved by raising the electron-accelerating voltage of the cathode beam in a sealed high-vacuum tube enough to provide a sufficiently bright single trace on the fluorescent screen for outside photography and recording.[5] This achievement eliminated the need for all pumping equipment and inside-the-vacuum photography, resulting in substantial simplification of the high-speed oscillograph and a further widening of the field of its application.

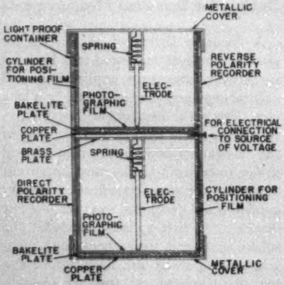


Fig. 2. Double-polarity surge voltage recorder for the RC voltage-time recorder

The cathode-ray oscillograph of high writing speed is capable of recording with adequate accuracy either timed or untimed transients or voltages, both as to amplitudes and wave shapes. The oscillograph itself is simple, but the auxiliaries required for time axis control, transient timing, and photographic recording of random circuit transients, are somewhat elaborate and expensive. As transients in electric circuits invariably involve wide ranges of amplitude and time, a study of random surges must be done on a statistical basis including many or at least several oscillographs and many records. The complete recording of a single transient disturbance at a number of points in a circuit must be done over and over again until the ranges of amplitudes and wave shapes have been successfully recorded. Such measurement studies are thorough and comprehensive but expensive. It was early recognized that much could be gained if low-cost crest voltage or crest current instruments recording only amplitudes could be placed at many measurement stations in a circuit. Fortunately therefore, as the new oscillographs were being more widely applied, several very useful low-cost recording devices were introduced to supplement the complete wave shape recording of the cathode-ray oscillograph and to reduce the necessity of applying it at so many points in the circuit.

Surge Voltage Recorder

In 1924, Peters suggested the application of photographic Lichtenberg figures to the measurement of surge voltages, and described the Klydonograph.[6] In 1926, the surge voltage recorder made available the double-polarity Lichtenberg figure recorder and the measurement of all crest voltages with positive polarity Lichtenberg figures. Also practical methods of measuring surge voltages on transmission lines were described, and lightning voltages on line conductors reported.[7,8] Extensive application was made of Lichtenberg figures for power circuit transient voltage measurements during the succeeding years and crest values and attenuations were determined by statistical analysis of thousands of records. The surge voltage recorder is still the only low-cost, automatic recorder capable of registering amplitudes and times of occurrence for surge voltages of random occurrence.

Flashover Indicator

As records of surge voltages accumulated thorough use of early recorders, the need for a simple surge voltage indicator

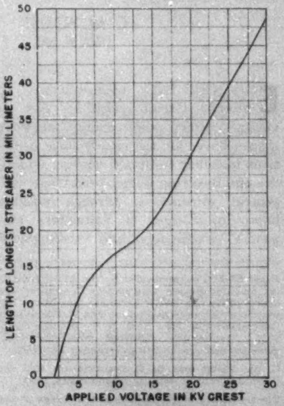


Fig. 3. Positive-polarity Lichtenberg figure calibration

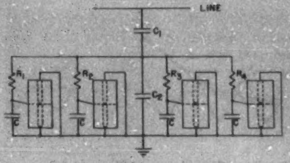


Fig. 4. Diagrammatic arrangement of 4-element RC recorder components

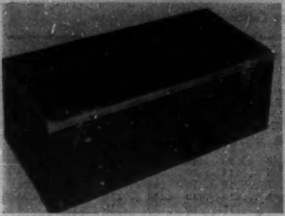


Fig. 5 (above). RC voltage-time recorder ready for installation in field

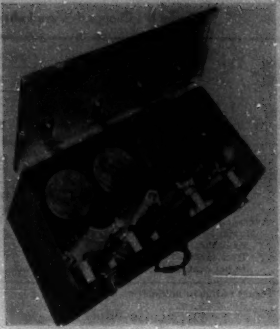


Fig. 6 (right). RC voltage-time recorder showing double-polarity film holders at top and RC elements at bottom

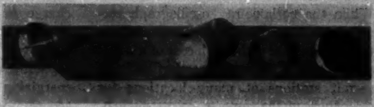


Fig. 7. Double-polarity voltage recorder removed and disassembled for film loading

became quite apparent. The particular requirement was for an indicating device which, when connected to a conductor or a piece of grounded structure, showed that a surge voltage had passed through it. One specific case was the arm of a tower when an insulator assembly on a transmission line flashed over. These requirements were met by a device later known as the flashover indicator.[9,10] This indicator contained, as the sensitive element, a frangible link which disrupted explosively when sparked through by a low surge voltage. This link disruption tripped into view a yellow-ocher disk easily visible from a distance indicating that a surge voltage had passed through the member to which it was connected. Many single-link indicators were installed on tower arms and good statistical surveys were made of the extent of insulator flashovers with lightning strokes of widely ranging amplitudes and points of contact with the line and its supporting structure. This sensitive, explosive, frangible link indicator was not appropriate for measurement of voltage or current of any kind, nor was it so used.

TRANSIENT ANALYZER

The responses of transmission, conversion, or utilization circuits to electric transients is quite different from that for the steady-state excitations for which they are designed. Since all electric circuits are inadvertently subjected to transient excitations and since these may constitute hazards to steady-state excitation, it is important that transient response be reasonably well understood. This applies particularly to transient voltage and insulation stresses, and to this end there was introduced in 1936 a transient analyzer.[11] This device measures transient voltage responses of circuits by applying selectable surge wave shapes and measuring the resulting voltages at required places in the circuit. Use of the transient analyzer supplemented the measurements of voltage surges under operating conditions by making available a well-controlled measurement analysis instrument capable of demonstrating experimentally the transient mechanisms required. The close relation of this device with other surge measurement devices for random transient measurement is therefore based on the fact that measurements with both have the same over-all objective: the determination of transient exposures and their effective control to avoid dielectric hazards.

SURGE CREST AMMETER

Early measurements of transient voltages and voltage wave shapes provided a substantial technical basis for conclusions as to insulation failures due to overvoltages. However, the early concentration on direct voltage measurements also showed the importance of transient current values and the need for a low-cost crest current recorder to be used on a many-measurement station basis and to obtain crest current values to correlate with crest voltage values from the surge voltage recorder. The important specific need which initiated the development of the surge crest ammeter was for measurements of lightning currents on transmission lines.[12,13]

Within the next few years, thousands of surge current crest measurements were made on many kinds of circuits. The surge crest ammeter proved particularly adapted to lightning currents on transmission conductors and in transmission line grounded structures. By installing measurement stations at many locations on a conductor or grounded structure network, the appropriate point of entry of the transient could now be readily determined, and its spread throughout the network recorded.[14,15]

New Voltage-Time Recorder

The developments and applications which produced and proved the indicators, instruments, and recorders described in the foregoing made available measurement devices and methods quite adequate for experimental study of many transient hazards. As information has increased of late, however, the lack of knowledge of time durations and wave

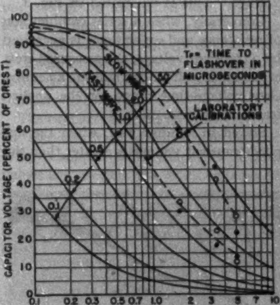


Fig. 8. Voltage-time calibration of recorder using chopped waves

Table I. Chopped-Wave Laboratory Calibration, Slow μsec Wave

	Resistance-Capacitance, μsec											
	0.1			1.73			3.3			5.0		
	FR, Mm*	Kv	Ratio, Per Cent	FR, Mm	Kv	Ratio, Per Cent	FR, Mm	Kv	Ratio, Per Cent	FR, Mm	Kv	Ratio, Per Cent
	46	28.5	95	26	18	60	20	14	47	13	6.5	22
	47	29	97	28	19	63	18	12	40	17	10	33
	48	29.5	98	24	17	57	19	13	43	14	7	24
	46	28.5	95	23	16	53	19	13	43	12	6	20
	47	29	97	23	16	53	17	10	33	18	12	40
Average	46.8	28.9	96.5	24.8	17.2	57.2	18.6	12.4	41.4	14.8	8.0	27
Maximum deviation from mean in per cent			1.55			10			20			48

* Figure radius in millimeters.

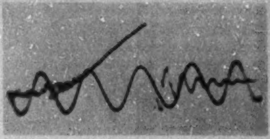


Fig. 10. Cathode-ray oscillogram of chopped wave of fast front; 1-megacycle timing wave superimposed

shapes has become quite troublesome. While the cathode-ray oscillograph records both amplitudes and time variations with good fidelity, the high cost of a single automatic measurement station will not permit stations at many locations for recordings of the same transient disturbance. The crest recorders such as the surge voltage recorder and surge crest ammeter give no information on the times involved for a particular crest recording. With all these available therefore there was still left an urgent need for a low-cost crest voltage recorder which would also record the time durations of surges of random time of occurrence. The *RC* (resistance-capacitance) voltage-time recorder was developed to fill this need.

Single-Unit RC Recorder

If a transient voltage energizes a resistance and capacitance in series, i.e., a specific *RC* combination, the crest voltage of the capacitor will be determined by the voltage-time characteristic of that transient, up to the time of crest capacitor voltage. This is illustrated in Fig. 1.

Conversely, when a series combination of resistance and capacitance is energized by a transient voltage, the measured crest capacitance voltage can indicate an equivalent voltage-time range for the applied transient voltage.

Stated in another way, when the applied transient voltage is greater than the capacitor voltage, an increasing capacitor charge and voltage will accumulate in proportion to the excess of transient voltage over the capacitor voltage. The capacitor voltage therefore increases with energizing transient voltage and time, in accordance with the specific values of the resistance and the capacitance in the combination. This *RC* combination therefore has an inherent time characteristic which is specific and quantitative and can be used as a measure of time or equivalent time when the method of its use is defined. A good mathematical illustration of this quantitative relation is the case of a square voltage wave front V applied to the series connected resistance and capacitance. The capacitor voltage

$$V_c = V\left(1-\epsilon^{-\frac{t}{RC}}\right)$$

will reach $(1-1/\epsilon)$ volts when t is *RC* microseconds.

A simplified and convenient utilization of the relation between capacitor voltage and the voltage-time content of a specific transient voltage energizing an *RC* combination can be based on the ratio of the crest voltage of the capacitor V_c to the crest voltage of the applied transient V. If this ratio V_c/V can be recorded, it will, when taken together with the *RC* which provided it, indicate a defined equivalent time content for the applied transient voltage. Presume, for instance, that a ratio of 0.5 is defined as the indicating ratio. Then the *RC* which gives a V_c/V of 0.5 would be accepted as the equivalent time content of the applied transient. If a smaller *RC* were to give a ratio approaching 1.0 this *RC* would be accepted as being faster than the applied transient. Again, if a larger *RC* were to give a ratio of 0.1 then this *RC* in microseconds would be accepted as being slower than the applied transient. Obviously, if applied waves were repeated and V_c/V ratios measured for many *RC* values the time order, and much detail concerning the equivalent voltage-time wave shape could be interpreted. Also, if many *RC* combinations were energized by one transient voltage, and a V_c/V measured for each *RC*, the same information would be obtained. With some modifications this is the basis for the *RC* voltage-time recorder described in this paper.

Multiple-Unit RC Recorder

The most important specifications for a voltage transient instrument capable of recording voltage and time for single transients of unknown time of occurrence are:

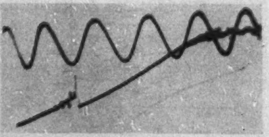


Fig. 9. Cathode-ray oscillogram of chopped wave of slow front; 1-megacycle timing wave superimposed

Table II. Chopped-Wave Laboratory Calibration, Fast μsec Wave

	Resistance-Capacitance, μsec											
	0.1			1.73			3.3			5.0		
	FR, Mm*	Kv	Ratio, Per Cent	FR, Mm	Kv	Ratio, Per Cent	FR, Mm	Kv	Ratio, Per Cent	FR, Mm	Kv	Ratio, Per Cent
	43	27	90	16	9.0	30	11	5.5	18	7	4.0	13
	39	25	83	20	14.0	7	14	7.0	23	10	5.0	17
	46	28.5	95	14	7.0	23	14	7.0	23	5	3.0	10
	48	29.5	98	15	8.0	27	17	12.0	40	3	2.7	9
	44	27.5	92	18	12	40	10	5.0	17	4	3.0	10
Average	44	27.5	92	16.6	10	33	13.2	6.9	23	5.8	3.5	12
Maximum deviation from mean in per cent			10			42.5			74			42

* Figure Radius in millimeters.

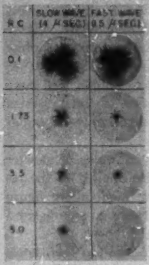


Fig. 11. Lichtenberg figures from chopped wave calibration

1. Automatic crest voltage registration.
2. Unlimited registration speed.
3. Recording energy below transient distortion.
4. Several decades of time range coverage.
5. At least 10-to-1 voltage range.
6. Low cost per recording station.

Specifications 1, 2, and 3 must be met by such means as are adopted to record crest voltages or voltage-time accumulations. To meet such requirements the registration means must have no inertia lag even though the time elements involved are in the order of hundredths of a microsecond. To meet specification 4 for generalized transients, such as in lightning and switching, the time ranges required include the decades 0.01, 0.10, 1.0, 10 and 100 μsec (microseconds). Meeting specification 5 presumes that low-energy capacitance voltage dividers can be used with the direct 10-to-1 recorder voltage range to extend the crest voltages covered up to several million when required. To meet specification 6 a single recording station cost must not be more than $200 or $300, even for the top of the lightning transient voltage range.

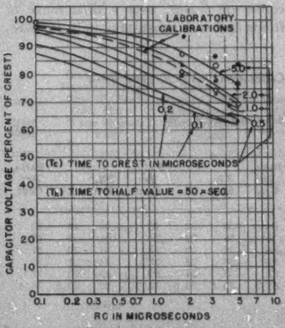


Fig. 12. Voltage-time calibrations of RC recorder using full waves

Method of Crest Voltage Measurement

Wide experience with the double-polarity Lichtenberg figure type of surge voltage recorder has proved its adequacy for crest voltage recording of transients of unknown time of occurrence. The double-polarity stationary-film type of surge voltage recorder was therefore selected to measure the crests of capacitor voltages in the *RC* combinations. The compact design shown in Fig. 2 was arranged to provide maximum flashover voltage with a 10-to-1 registration ratio, 2.5 to 25 kv, and to fit into a multiple-recorder combination.

A crest voltage-figure size calibration was determined by applying voltage transients from registration to flashover voltage. These direct calibrations of the recorder showed this relation to be as shown in Fig. 3.

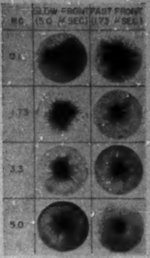
This calibration was made on the recorder arrangement shown in Fig. 2. It shows the total recording voltage range to be from 2 to 30 kv, which is from the corona threshold to recorder flashover. Corresponding Lichtenberg figure radii range from a few millimeters to 50 millimeters. This calibration curve was made with full waves of 50 μsec duration and checked with chopped waves of time durations less than 0.1 μsec. Figure size therefore as measured by the length of the longest streamer and for positive Lichtenberg figures does not vary appreciably with wave duration.[16]

Complete RC Voltage-Time Recorder

While each of the five decades, 0.01 to 100 μsec, may be of concern in some specific circuit transient cases, experience has indicated that for any one recorder station on a particular circuit a selection of four decades would suffice. For the shorter, faster circuits the first four, 0.01, 0.10, 1.0 and 10 μsec, would be most appropriate, while for longer, slower circuits the last four, or, 0.10, 1.0, 10.0 and 100 μsec, would be adequate. Accordingly, the 4-element recorder shown diagrammatically in Fig. 4 was made the basis for a general-purpose recorder.

Such a 4-element recorder may be readily adjusted to cover four selected decades or any four specific *RC* time equivalents as required. All the elements can be assembled in one housing as shown in Figs. 5, 6, and 7 with only two outside connections. The metal housing will normally be connected to ground and the opposite terminal to the source of transient voltage, a conductor or an antenna inductively coupled to the conductor.

Fig. 13. Lichtenberg figures from full-wave calibration



Recorder Calibrations

Calculations: Chopped Waves

An important application of the *RC* voltage-time recorder is that of insulator assembly flashover on power lines. A few cathode-ray oscillograms of such flashovers indicate the flashover to be on the front of a rising voltage wave which then is chopped to practically zero voltage.[17] In Fig. 5 of reference 17 is an oscillogram showing the major portion of the flashover voltage rise occurring within a 2-μsec period over a voltage range of 2 million, giving a rate of rise of more than 1 million volts per microsecond. Acceptance of such a uniformly rising voltage as one significant wave shape for the calibration of the *RC* voltage-time recorder is therefore highly warranted.

Based on the total direct recording voltage range of 2 to 30 kv, a suitable practical measurement range is 2.5 to 25 kv, which is the 10-to-1 range required. For good utilization of this range on a high-voltage line, an insulator assembly flashover voltage of 2,500 kv might well be associated with the top recording range of 25 kv. This sets up the requirement for voltage divider ratio of 2,500/25 or 100 to 1. Practical divider ratios for other line voltages are similarly calculated.

The calibration of the recorder can now be considered as a direct calculation. When chopped wave shapes of several values are selected, the relations between these and the ratios of crest recorder voltage to crest capacitor voltage are readily calculated from the following relation:

$$V_c = S\left[T_f + RC\left(\epsilon^{\frac{-T_f}{RC}} - 1\right)\right]$$ (see the Appendix for derivation)

V_c = capacitor voltage

S = slope of voltage across recorder

T_f = time to flashover

Table III. Full-Wave Laboratory Calibration, 5x50 μsec Wave

	Resistance-Capacitance, μsec											
	0.1			1.73			3.3			5.0		
	FR, Mm*	Kv	Ratio, Per Cent	FR, Mm	Kv	Ratio, Per Cent	FR, Mm	Kv	Ratio, Per Cent	FR, Mm	Kv	Ratio, Per Cent
	47	29	97	37	24	80	36	23	77	31	20	67
	48	30	100	41	26	87	38	24	80	35	23	77
	48	30	100	41	26	87	40	25	83	36	23	77
	49	30	100	43	27	90	41	26	87	37	24	80
	49	30	100	46	28	93	41	26	87	40	25	83
Average	48.2	30	99.4	41.6	26.2	87	33.3	24.8	83	35.4	22.8	76
Maximum deviation from mean in per cent			3.3			8.4			7.3			12

* Figure radius in millimeters.

R = resistance of RC combination

C = capacitance of RC combination

ε = base of natural logarithms

Since the most necessary quantitative information is that of time to flashover, a calculated plot is shown in Fig. 8 of the ratio of measured voltages against RC time constants for six specific times to flashover ranging from 0.1 to 5 μsec.

With such a pattern of relations available and assuming a uniformly rising voltage wave, the time to crest flashover voltage can be read directly at the several values of RC incorporated in the recorder. Obviously, only such ordinate values of voltage ratios will be available as are within the recording range of the Lichtenberg figures registered. If it is required that several voltage ratios be recorded for check purposes then it is only necessary that RC values selected for the recorder be sufficiently close to provide such readings.

Laboratory Calibration: Chopped Wave Applied

For the purpose of demonstrating a typical utilization, one of the recorders was equipped with four RC values 0.1, 1.73, 3.3, and 5.0 μsec. This spacing of time constants was selected to provide a considerable range of time value, and still ensure that two or more recorders would be checking each other on time to crest of the applied wave. A surge voltage generator was connected for chopped front output and a slow rate of voltage rise. A single wave was applied to the RC recorder giving four positive polarity Lichtenberg figures which, when referred to the calibration of Fig. 3, gave four corresponding capacitor voltages, one for each RC combination. These values are shown in Table I on the top horizontal line, along with their corresponding voltage ratios based on an applied 30-kv crest.

The ratios of line 1 in Table I have been plotted in Fig. 8 in dots. An indicated time to crest of about 4 μsec is well checked by all four RC points. To determine the spread of indicated times to crest, four repeat applications of the same wave were applied. Lines 2, 3, 4 and 5 record the voltages and ratios obtained. These have been averaged and are plotted in Fig. 8 in circles. The average points as expected are closer to the 4-μsec time to crest than the single figure points of the first surge application.

The maximum deviations of a single record point from the average of 5 are shown on the bottom line of Table I. These percentage deviations range from a maximum of 38% to 1.8%. The deviation of the time to crest finding for an unknown wave would obviously be within these percentages and might well be taken to be 10%. In fact, in this case, the 4 μsec to crest from the RC recorder was in close agreement with the time measured on the cathode-ray oscillogram of the applied wave as shown in Fig. 9.

In this oscillogram attention should be given to the superimposed oscillation at about mid-voltage. The RC recorder simply integrated voltage and time through this oscillation, finally indicating the correct over-all equivalent voltage and time as applied up to chopping. Any measurement method sensitive to voltage or current slopes might well have given a false indication of time to crest because of such a superimposed oscillation.

A similarly chopped wave of shorter time to crest was also applied to the recorder with resulting Lichtenberg figure sizes, voltages, and ratios as summarized in Table II.

Again, in Table II, the top line shows the four recordings and derived ratios for the first application of the fast wave. These four points are also plotted on Fig. 8 in dots, indicating a wave of about 1.5 μsec to the chopping point at the crest. Four more applications of this wave were made and are included in Table II. Average ratios are also shown and plotted in circles in Fig. 8. The calculated percent maximum deviations at the bottom of Table II range from 10% to 53%. The cathode-ray oscillogram of this chopped wave of fast front indicated 1.75 μsec to crest as shown in Fig. 10.

Two sets of Lichtenberg figures, one for slow front and one for fast front, are reproduced in Fig. 11. The higher voltages for the slower waves are clearly evident in the larger figures at the left.

Calculations: Full Waves

At the other end of the time content scale from the chopped wave is the full-wave transient, i.e., a wave having a definite front or rising portion and a substantial tail or falling portion. Mathematically, such a wave can often be represented as the sum of two exponentials. Full-wave transients are quite common on power and industrial circuits from lightning, switching, and other excitation disturbances. A calibration of the RC recorder on waves, over a time range of fronts and with substantial tails, is there-

Table IV. Full-Wave Laboratory Calibration, 1.73x50 μsec Wave

	Resistance-Capacitance, μsec											
	0.1			1.73			3.3			5.0		
	FR, Mm*	Kv	Ratio, Per Cent	FR, Mm	Kv	Ratio, Per Cent	FR, Mm	Kv	Ratio, Per Cent	FR, Mm	Kv	Ratio, Per Cent
	47	29	97	37	24	80	33	22	73	29	19	63
	47	29	97	37	24	80	34	22	73	31	20	67
	48	30	100	39	25	83	34	22	73	31	20	67
	48	30	100	39	25	83	35	23	77	34	22	73
	49	30	100	40	25	83	38	24	80	35	23	77
Average	47.8	29.6	98.8	38.4	24.6	81.8	34.8	22.6	75.2	32	20.8	69.4
Maximum deviation from mean in per cent			1.8			2.2			6.4			11

* Figure radius in millimeters.

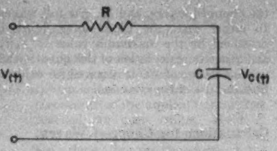
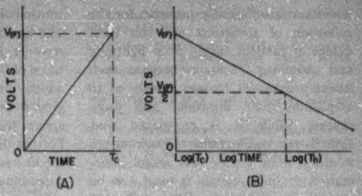


Fig. 14 (left). The RC circuit

Fig. 16 (right). The full wave

A—Front of wave
B—Decrement of wave



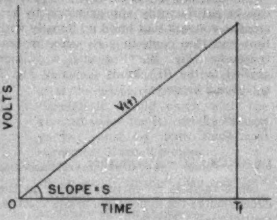


Fig. 15. The chopped wave

fore necessary. For the purpose of showing wave front effects as compared to the chopped wave calibration of Fig. 8, the same range of fronts from 0.1 to 5 μsec was retained but each front was now given a 50-μsec to half-crest value tail. Corresponding values of *RC* in microseconds and ratios of applied-to-capacitor voltages were calculated for the six full waves selected. This relation is plotted in Fig. 12. The method of calculation is given in the Appendix. Comparing Fig. 12 with Fig. 8, both of which are on the same co-ordinate scale, clearly shows the substantial time content of the 50-μsec tail. Nevertheless, each of the six wave fronts is distinctly separable in the *RC* voltage-time relation.

LABORATORY CALIBRATION: FULL WAVE APPLIED

To demonstrate the utilization of the recorder on full waves a surge voltage generator was arranged to provide 50-μsec wave tails for each of the six fronts ranging from 0.1 to 5.0 μsec.

Some of the data obtained on the 5.0x-50-μ sec wave are summarized in Table III. In this table the figure sizes and voltages are arranged vertically in increasing order of magnitude to show the range of variations more readily. The minimum and maximum ratios for corresponding *RC* values are plotted in Fig. 12 in dots, thus showing the maximum spread of indicated time to crest for five surges applied. The average of the five surges is plotted in circles showing the wave time to crest to have been 5 microseconds by the cathode-ray oscillogram.

Some of the data on a 1.73x-50-μsec wave are summarized in Table IV, which is arranged the same as Table III and plotted in Fig. 12 slightly to the left of the slow-wave data. The indicated time to crest is maximum at 2 μsec and minimum at 1 μsec with a suggested value a little over 1.5 μsec.

It is important to note that the distinction between the 5-μsec time to crest and the 1.73 μsec time to crest is very decided even with the 50-μsec wave tail. It seems clear that the *RC* method with Lichtenberg figure recording of capacitor crest voltages will differentiate well between any of the calculated time to crest values as calculated and plotted in Fig. 12 for the long-tailed wave.

The Lichtenberg figures shown in Fig. 13 are selected to be representative of the 5x50-μsec and 1.73x50-μsec full-wave calibrations. Comparison of this figure with Fig. 11 for the chopped waves shows how readily apparent the time contents of the full against the chopped waves and the fast fronts against the slow are made with these corona splashes on photographic film.

Generalized Interpretations

While Lichtenberg figures are not adapted to the measurement of crest transient voltages on a precision basis[7] they are

Table V. Method for Calculating Capacitor Voltage When Full Waves are Applied

Column Designation:	A	B	D	F	G	H	J	K	M
Row No.	Instantaneous Impulse Voltage Applied (Read from Plot of Wave Shape), $V(t)$	Average Impulse Voltage in Incremental Time Interval, $V(t)_{avg}$	Instantaneous Capacitor Voltage $V_c(t)$	Average Capacitor Voltage in Incremental Time Interval, $V_c(t)_{avg}$	Average Voltage Across Resistance in Incremental Time Interval, $V(t)_{avg}-V_c(t)_{avg}$	Instantaneous Time t	Incremental Time Interval Δt	Increment in Capacitor Voltage in Incremental Time Interval = $\frac{V(t)_{avg}-V_c(t)_{avg}\Delta t}{2CR}$ $V_C(t)$	Capacitance Voltage at End of this Incremental Time Interval (Should Check with Value Previously Assumed for this Quantity in Column D)
1	$V(F)$		$V_C(F)$ (calculated)			T_C			
2		$\frac{A_1+A_3}{2}$		$\frac{D_1+D_3}{2}$	B_2-F_2		Δt	$\frac{G_2J_2}{2CR}$	D_1+K_2 (should equal D_3)
3	$V(\Delta t)$		$V_C(\Delta t)$, assumed			$T_C+\Delta t$			
4		$\frac{A_3+A_5}{2}$		$\frac{D_3+D_5}{2}$	B_4-F_4		Δt	$\frac{G_4J_4}{2CR}$	D_3+K_4 (should equal D_5)
5	$V(2\Delta t)$		$V_C(2\Delta t)$, assumed			$T_C+2\Delta t$			

. . . Continued until the capacitance voltage (column M) reaches a maximum value.

nevertheless sufficiently accurate for the collection of statistical data on crest values of random surges from lightning and switching. This is because such surges are known to vary widely as to amplitude, time duration, and wave shapes. While it is recognized that cathode-ray oscillograms of such surges would provide precise and useful information on single surges, it must also be recognized that the small number of oscillograms which could be obtained at a tolerable cost would not suffice to portray amply the range and average values required. Only many records and their statistical analysis will provide this, as has been done for crest values of surge voltages and surge currents with the surge voltage recorder and surge crest ammeter. Many *RC* voltage-time recorder stations are required for long-range system studies.

The calibrations described have been on impulse waves, i.e., single-polarity transients. Obviously, surges of oscillatory shapes and repeated surges of mixed polarity will be present on many electric circuits and will be impressed upon the instruments connected for their measurement. Fortunately, the Lichtenberg figures will provide some information for analysis of such surges.

Because each Lichtenberg figure represents one single-polarity voltage surge and because overlapping Lichtenberg figures may be easily recognized, it is usually possible to interpret composite figures as representing several surges. Also, with the double-polarity recorder when oscillatory surges or repeated surges of mixed polarity are present, the presence of superimposed positive- and negative-polarity Lichtenberg figures on both the directly connected and reversed recorders will permit considerable analysis. Some experience in recognizing polarity characteristics and in noting the crossed streamers of repeated impulses aids greatly in reaching adequate interpretations.

Since the direct measurement range of the recorder is 2.5 to 25 kv, it is necessary to use low-capacitance voltage dividers such as have been extensively used with the surge voltage recorder to apply the new instrument to higher voltage ranges and circuit ratings. The highest value of normal circuit excitation permissible across the recorder is 2.5-kv crest, at which the divider ratio for a 3-phase circuit rating for conductor to ground measurements would be

$$\frac{\text{kv circuit rating}}{\sqrt{3}} \times \frac{\sqrt{2}}{2.5}$$

an example being 75 to 1 for a 220-kv transmission circuit, or 4.5 to 1 for a 13.8-kv distribution circuit. Since the recorder has a built-in parallel capacitance of 0.001 μf (microfarad) to which is added other parallel capacitance to ground of 0.0005 microfarad, the total across the recorder will be 0.0015 μf. For the divider ratio of 75 to 1 therefore the line side capacitance required will be approximately $0.0015/75 \times 10^{-6}$ or 20 $\mu\mu f$, and for the 4.5-to-1 ratio 333 $\mu\mu f$. These values are illustrative only. The low-energy requirement and low capacitance of the recorder make the problem of voltage divider a simple one, although it must be recognized that when such a divider is energized at high voltages, all normal voltage distributions are rearranged by the extending streamers of a possible flashover in the making.

Conclusions

1. In response to the need for a low-cost transient voltage-time recorder for random surge voltage measurements, the *RC* voltage-time recorder has been made available.
2. The *RC* recorder is added to the already available line of crest voltage and current recorders and indicators and includes for the first time the requirement of time-to-crest measurement.
3. The low-energy requirements and low capacitance of the new recorder make its connection to power lines of any voltage through a capacitance divider a simple one.
4. Calculated calibration curves for the recorder covering four decades of time interval are shown to agree with laboratory tests using surge generator wave shapes.
5. The Lichtenberg figure method of recording crest voltages is shown to be appropriate to the ranges and types of surges to be measured.

Appendix. Methods of Calculation of Crest Capacitor Voltages

In the linear *RC* circuit shown in Fig. 14:

$V(t)$ = applied voltage as function of time
R = resistance in circuit, ohms
C = capacitance in circuit, farads
$V_c(t)$ = voltage across capacitor as function of time

$$\text{Then } V(t) = Ri + \frac{1}{C}\int_0^t i\, dt + V_c(0) \qquad (1)$$

where $V_c(0)$ = capacitor voltage at beginning of $V(t)$, i.e., at $t=0$
i = instantaneous current in circuit, amperes

$$\text{i.e., } \frac{dV(t)}{dt} = R\frac{di}{dt} + \frac{i}{C} \qquad (2)$$

The voltage-time recorder is designed for recording lightning wave magnitudes and wave shapes. In the test the lightning waves have been divided in two classes: 1. the chopped waves, and 2. the full waves. In either case, the Lichtenberg figure is produced by the maximum value of $V_c(t)$, and hence the calculation of this quantity for various values of *RC* is required for calibration and use of the recorder.

Calculations for Chopped Waves

The chopped waves observed in practice can be satisfactorily approximated by triangular voltages that build up linearly with time and then come to zero value instantaneously (Fig. 2). For such a voltage applied to the *RC* circuit shown in Fig. 1 equation 2 becomes

$$S = R\frac{di}{dt} + \frac{i}{C}$$

$$t = -RC \ln(SC - i) + \text{constant}$$

Since there is no current flowing in the circuit before the application of the voltage, $i=0$ at $t=0$. Hence

$$\text{constant} = RC \ln SC$$

$$i = SC\left(1 - \epsilon^{-\frac{t}{RC}}\right) \qquad (3)$$

The maximum value of voltage across the capacitor

$$V_c \text{ peak} = \frac{1}{C}\int_0^{T_f} i\,dt + V_c(0)$$

where $V_c(0)=0$ since there is no charge on the capacitor at the time of application of the surge voltage.

$$V_c \text{ peak} = S\int_0^{T_f}\left(1 - \epsilon^{-\frac{t}{RC}}\right)dt$$

$$= S\left[T_f + RC\left(\epsilon^{-\frac{T_f}{RC}} - 1\right)\right] \qquad (4)$$

Fig. 15 shows the chopped wave.

Calibrations for Full Waves

Available data on wave shapes of surge voltages on transmission lines indicate the existence of a wide variation in these, and hence an equivalent wave shape in terms of which the results obtained with the use of the voltage-time recorder will be interpreted, is defined in Fig. 16(A) and (B). This equivalent wave shape is supposed to come up to the crest value with a straight-line voltage-time relation as shown in Fig. 16(A) and it is assumed to decay as shown in Fig. 16(B) where the voltage versus logarithm (time from start of wave − time to crest of wave) is a straight-line relation. Then the wave shape is identified by the time to crest T_c, and the time to half-value T_h.

When such a voltage is applied to the *RC* circuit of Fig. 14, the capacitor voltage cannot be easily calculated analytically from equation 2, and hence the results were obtained by the following method. First, the voltage across the capacitor V_c was calculated at the time when the surge voltage reached the crest value. This calculation is identical to the calculation of the maximum value of capacitor voltage for a triangular voltage applied to the circuit. This voltage known, calculations for the capacitance voltage were continued for the decaying portion of the surge voltage wave by a step-by-step method shown as follows: Let

$V_c(F)$ = capacitor voltage when the applied voltage reaches its crest value

$V_c(K\Delta t)$ = capacitor voltage at a time $K\Delta t$ seconds after the applied voltage has reached its crest value. Here K is a constant having values 1, 2, 3, etc., depending on the stage of the calculation. For the first step, when $V_c(F)$ is known, and the incremental time (Δt) to be considered has been decided on, the value for $V_c(\Delta t)$ must be guessed. Then, for the value calculated for $V_c(\Delta t)$ to be correct, after the calculations have been gone through it should check with the guessed value for it; otherwise new guesses must be made until such agreement of result is obtained.

Δt = small increment of time. This is chosen to be much smaller than the time constant of the RC circuit for accuracy of calculations

$V(F)$ = crest value of the voltage applied to the RC circuit

$V(K\Delta t)$ = magnitude of the applied voltage $K\Delta t$ seconds after it has reached its crest value. Here $K = 1, 2, 3$ similarly to that in $V_c(K\Delta t)$

Then the average value of the current flowing through this circuit during the interval Δt after the voltage has reached its crest value is

$$i_{avg} = \frac{\dfrac{V(F)-V_c(F)}{R} + \dfrac{V(\Delta t)-V_c(\Delta t)}{R}}{2}$$

$$= \frac{\dfrac{V(F)+V(\Delta t)}{2} - \dfrac{V_c(F)+V(\Delta t)}{2}}{R}$$

and the increase in the capacitance voltage at the end of this time interval is the following equation

$$\Delta V_c(\Delta t) = \frac{1}{C} i_{avg} \Delta t$$

$$= \frac{(V(t)_{avg} - V_c(t)_{avg})}{2CR} \Delta t$$

Now

$$V(F) + \frac{i_{avg}(\Delta t)}{C} = V_c(\Delta t)$$

This calculated value of $V_c(\Delta t)$, for the calculations to be correct, should check with the value guessed for it at the beginning of the calculations. Such calculations were continued until $V_c(t)$ reached its peak value. The form used for these calculations is shown in Table V.

References

1. The Sphere Spark Gap, S. W. Farnsworth, C. L. Fortescue. *AIEE Transactions*, vol. 32, pt. I, 1913, pp. 733–37.

2. Calibration of Sphere Gap Voltmeter, L. W. Chubb, C. Fortescue. *Ibid.*, pp. 739–48; *discussion* by F. W. Peek, Jr., pp. 812–20.

3. Measurement of Voltage in Dielectric Tests. *ASA Standard C68.1*, American Standards Association, New York, N. Y., 1953.

4. L'onde Electrique, Vol. 1 (book), A. DuFour, 1922, pp. 638, 699.

5. A New High-Speed Cathode-Ray Oscillograph, H. P. Kuehni, Simon Ramo. *AIEE Transactions*, vol. 56, June 1937, pp. 721–28.

6. The Klydonograph, J. F. Peters. *Electrical World*, New York, N. Y., vol. 183, Apr. 19, 1924, pp. 769–73.

7. The Measurement of Surge Voltages on Transmission Lines Due to Lightning, Everett S. Lee, C. M. Foust. *AIEE Transactions*, vol. 46, 1927, pp. 339–48.

8. Symposium on Surge Voltage Investigations, W. W. Lewis. *Ibid.*, vol. 47, Oct. 1928, pp. 1111–21.

9. Instruments for Lightning Measurements, C. M. Foust. *General Electric Review*, Schenectady, N. Y., Apr. 1931.

10. Lightning Investigation, Edgar Bell, A. L. Price. *AIEE Transactions*, vol. 50, Sept. 1931, pp. 1101–10.

11. The Oscillograph Electric Transient Analyzer, N. Rohats. *General Electric Review*, Mar. 1936.

12. The Surge Crest Ammeter, C. M. Foust, H. P. Kuehni. *Ibid.*, Dec. 1932, p. 644.

13. Direct Measurement of Surge Currents, C. M. Foust, J. T. Henderson. *AIEE Transactions* (*Electrical Engineering*), vol. 54, Apr. 1935, pp. 373–78.

14. Lightning Investigation on Transmission Lines—VI, W. W. Lewis, C. M. Foust. *Ibid.*, vol. 56, Jan. 1937, pp. 101–06.

15. Lightning Stroke Protection at High Altitude in Peru, C. M. Foust, B. C. Maine, C. Lee. *Ibid.*, vol. 72, pt. III, June 1953, pp. 383–93.

16. Measurement of Transients by the Lichtenberg Figures, K. B. McEachron. *Ibid.*, vol. 45, 1926, pp. 712–20.

17. Lightning Investigation on Transmission Lines—IV, W. W. Lewis, C. M. Foust. *Ibid.*, vol. 53, Aug. 1934, pp. 1180–86.

◆

Discussion

J. H. Park (National Bureau of Standards, Washington, D. C.): The RC voltage-time recorder should make it possible to obtain statistical data on actual wave shapes of lightning voltages. Since this recording instrument offers a possibility of fulfilling a need which has not been met previously it should be applied in actual field measurements despite its drawbacks. These drawbacks, which may or may not prove to be serious, may be given as follows:

1. Both the computed and the experimentally determined calibrations of the recorder are predicted on specific wave shapes, i.e., either a linearly rising wave chopped on its rising front, or a wave with a linearly rising front which decays logarithmically with time to half value in 50 μsec. The actual wave shapes found on transmission lines may be so different as to cause the calibrations to give doubtful results.

2. One set of calibration curves is used for chopped waves (Fig. 8) and a different set for full waves (Fig. 12). Which set of curves to use for a given field record may not always be self-evident. For example, if a calibration curve for a 10-μsec-front chopped wave were added in Fig. 8, it would appear nearly the same as one of the full-wave calibration curves shown in Fig. 12.

3. A capacitance divider must be used for most measurements and, for steeply rising waves, the residual inductance of the high-voltage capacitor plus its connecting leads may introduce significant error.

4. The low-voltage capacitance of the divider circuit includes all the C's in the four RC circuits. Is C_1 in Fig. 4 large enough to swamp out the effects of these four RC circuits? Also, can residual inductance in these circuits be neglected?

Answers at this time are not expected to all these points, which are submitted mainly to emphasize that they must be considered in any application of the RC recorder.

H. Linck (Hydro-Electric Power Commission of Ontario, Toronto, Ont., Canada): The transient voltage-time recorder is an interesting and useful application of the principle of surge recording with Lichtenberg's figures. Attempts to determine surge wave shape with conventional klydonographs are often not successful, especially for complex patterns recorded in the field.

Although describing their instrument in general, the authors have not pointed out clearly enough that at least two units have to be used for the recording of one particular single surge. As an example, for the front-chopped wave equation 4, slightly modified to

$$V_c = V_f\left[1 + \frac{RC}{T_f}\left(\epsilon - \frac{T_f}{RC} - 1\right)\right]$$

shows that two unknown quantities, the surge crest voltage V_f and the time to chop T_f, have to be determined; this can be done only with two equations derived from two recording units with different time constants.

Assuming a linearly rising wave and with R_1C_1 and R_2C_2 being the time constants of the two klydonographs

$$V_f = \frac{V_{c1}}{p_1} = \frac{V_{c2}}{p_2}, \quad \text{where} \quad p_{1,2} = \frac{V_{c1,2}}{V_f} \times 100\%$$

$$\text{and} \quad \frac{p_1}{p_2} = \frac{V_{c1}}{V_{c2}}$$

From the calibration curves shown for the front-chopped surge (Fig. 8), V_f and T_f can be determined, for example:

$R_1C_1 = 0.2$ μ sec

$R_2C_2 = 1.0$ μ sec

$V_{c1} = 20$ kv

$V_{c2} = 10$ kv

$\frac{p_1}{p_2} = 2.0$

$p_1 = 82\%$

$p_2 = 41\%$

$V_f = 24.4$ kv

$T_f = 1.2$ μsec

As the authors mention, field records frequently are much more complex than patterns obtained in the laboratory; also considering the possible large recording error (up to 74% for full waves, according to Table II), it becomes obvious that for field records the measurement of front time in many cases may only result in an estimated order of magnitude. As the frequency of film changes is a matter of economics, especially for extended surveys, several surges may be recorded on one film, aggravating the difficulties of interpreting the records. But in spite of these drawbacks, the recorder represents the only practicable instrument for collecting statistical data on surge front durations and wave shapes.

An inexpensive double-polarity surge voltage recorder was developed 10 years ago by Ontario Hydro with a simplified coupling method, allowing installation without a separate coupling device at many locations and at any time.[1] This commercially available klydonograph, has been extensively used for transmission line, high-voltage station, and distribution system lightning and switching surveys, and it has rendered valuable service. Modification of this design to incorporate an *RC* element is possible, and the cost per unit would certainly be within the range considered acceptable by the authors.

REFERENCE

1. A Simplified Double-Film Klydonograph With an Improved Coupling Method, J. H. Waghorne. *AIEE Transactions*, vol. 66, 1947, pp. 1114–16.

C. M. Foust and B. V. Bhimani: The authors appreciate the comments of Mr. Park and Mr. Linck to the effect that the *RC* voltage-time recorder should make statistical data on actual wave shapes possible and that it provides the only practical instrument available.

That the *RC* recorder is suitable for application on high-voltage power lines has been determined by trial installations of a few recorders during the end period of the 1956 lightning season. With a few recorders so distributed that correlation of records could not be expected, the pattern of surge voltages was nevertheless clearly indicated, as including chopped-voltage waves, full waves and long-time transients of the switching surge type.

A chopped-wave record from one station gave voltage percentage points of 99, 68, and 56 for *RC* values of 0.1, 3.3, and 5.0 respectively. When plotted on a chopped-wave calibration, these three points all fitted on a 10-μsec-to-crest curve with good accuracy. The crest voltage recorded was 1,230 kv, which at 10 μsec gives a voltage steepness of 123 kv per μsec.

A full-voltage wave was also recorded, giving percentages of crest of 100, 67, and 56 for 0.1, 1.73, and 5.0-μsec *RC* elements, respectively. These points fit well on a 1.0-x10-μsec wave calibration. The crest voltage recorded was 480 kv, giving a voltage steepness of 480 kv per μsec.

In addition a considerable number of long-duration, conductor-positive, tower-negative surges ranging from crest line excitation voltage up to 480 kv were recorded. In all of these cases the total equivalent time of the surge was greater than the highest (5.0) *RC* of the installed recorder.

Mr. Park calls attention to possible drawbacks which may or may not prove serious. The first concerns the fact that the computed and experimentally determined calibrations presented in the paper are predicated on specific wave shapes, either one chopped on its rising front by flashover or a full wave including both rising front and slowly diminishing tail voltage. The authors agree that the actual voltage wave shapes found on power lines may be quite different. Review of the cathode-ray oscillograms obtained to date shows voltage wave shapes differing so widely that their classification into types, each having particular and important features, cannot be determined convincingly, at least until many more such oscillograms are available. Such oscillograms are so expensive to obtain that it now appears that such a typing of waves may never be possible. However, even the few field oscillograms now available and all laboratory flashover test oscillograms can readily be classified into one of two types, either the chopped wave of flashover or the full wave of no flashover, and this is the reason for the calibrations on these two wave types as calculated and checked experimentally in the laboratory and presented in the paper.

Mr. Park mentions a second possible drawback: that whether to use the chopped-wave or the full-wave curves for a given field record may not always be self-evident. It is true that we intentionally did not present detailed procedures for analysis of field data because field data suitable for trial analysis was not yet available and only laboratory data had been analyzed. Subsequently, the small amount of field data permitted trial analysis with very gratifying results in that no data were examined which left doubts as to its chopped- or full-wave origin. We have made quite extensive investigations of the curves of percentage crest voltages against *RC* equivalent times for various wave shapes and have found that, on a 4-point *RC* recorder basis, the possibility of both a chopped wave and a full wave of the same crest voltage value giving identical curves is very improbable. However, if field data of sufficient quantity should become available and if these should indicate such an uncertainty in interpretation, we would recommend measuring both *R* and *C* crest voltages, which could be readily done and which would clear up all doubts concerning chopped- or full-wave interpretations.

The possible drawback of the inductance of the high-voltage capacitor and its connecting leads is mentioned with good justification by Mr. Park. The rapid rates of local voltage changes involved in natural lightning and in laboratory surge voltage tests make this consideration a very important one.

We appreciate Mr. Park's contributions to the technical literature on this particular point and on other rapid rate-of-change problems. To guard against significant errors due to inductances in the *RC* recorder circuits, we have used the well-established method of calculating data from measured circuit constants and checking these calculations against data obtained from direct laboratory tests.

In reply to Mr. Park's question about the value of C_2 in Fig. 4, C_2 is large enough to "swamp out" the effects of the four *RC* circuits. The residual inductance of these circuits is made as low as possible and is then neglected because the comparison of calculated and measured percentage-voltage *RC* curves shows that such neglect is warranted.

The authors are happy to have Mr. Linck's observations, based on the contributions he and Mr. Waghorne have made to the problems of measurements of untimed surge voltages in power circuits. It is regretted that they omitted a direct statement on the need for two or more recorders of the *RC* type as necessary to wave slope measurements. The four units, incorporated in the recorder as described, serve this purpose, by obtaining four points on the "percentage voltage-*RC*" curve, as well as to extend the time range of the recorder so that all surges of voltage and time range pertinent to the objectives of the investigation are adequately recorded.

As regards the reproducibility of Lichtenberg figure diameters at one crest voltage, this is certainly a factor in the accuracy of any recorder using this method of recording crest voltages. It is true that for any one surge voltage the Lichtenberg figure recording of its crest value is not as accurate as desired. This has been well known throughout the three decades of their utilization and yet on a statistical basis these figures have proved the most practical source of information on crest voltages of uncertain surges. Such information has established the ranges and frequencies of occurrence of surge voltages on power circuits as well as the best information available on attenuation of crest voltage in distance along a transmission line.

Based on laboratory calibrations and sample field records it now appears that an extended and co-ordinated field installation of the *RC* recorder is warranted. It appears that wave front determinations will be adequate, especially if this installation is arranged to provide confirmation of results in several ways as follows:

1. Recorders on adjacent towers recording the same wave will supplement each other by showing in general a decrease in voltage steepness as the wave travels from the point of greatest steepness.
2. Correlating surge crest ammeter data will verify the tower location of direct stroke or highest current amplitude.
3. Flashover indications on insulators will supplement recorder slope indications.
4. Associated line tripouts will verify the occurrence of flashover.

The authors' conviction is strong that the long-felt need for an equivalent voltage wave front recorder of low cost and simplicity of operation will be well provided for by the *RC* recorder if a well-arranged installation is used and careful interpretation of records is maintained.

Transistor NOR Circuit Design

WILLIAM D. ROWE
AFFILIATE MEMBER AIEE

GEORGE H. ROYER
ASSOCIATE MEMBER AIEE

THE transistor NOR circuit was introduced because of the desire to have a single logic element which would combine well in logic configurations with a minimum of concern for matching and element loading. A NOR element has a signal output only if there are no input signals, and consequently has no output if any input signal is present. Such a single logic element must have a considerable amount of flexibility to meet the various input and output loadings, operating temperatures, and power-supply variations which may be encountered. This paper considers the circuit design of the transistor NOR circuit.

Fig. 1 shows a transistor NOR circuit with an arbitrary number of M inputs and loaded with N outputs. The equivalent output loading for each stage is a resistor R_1 to ground through the base impedance of a transistor and the total output impedance is R_1/N. V_{cc} is the collector supply voltage and the R_T to V_{BB} combination supplies a bias to reduce the transistor leakage current I_c to a minimum when the transistor is cut off. R_1 is the input-current-limiting resistor and also serves to prevent feedback of signals between inputs. The transistor operates in the switching mode, becoming saturated and offering a low conducting impedance when on, and, when blocking or off, allowing only a small leakage current to flow. The polarities of the supplies shown in the circuit are for a p-n-p transistor, yielding negative voltage signals throughout the system. The polarities could be reversed and n-p-n transistors substituted, with a corresponding change to positive voltage signals throughout the logic system. P-n-p and n-p-n transistor NOR circuits as shown cannot be readily intermixed, however.

Transistor Test Methods

In designing a NOR circuit for a particular manufacturer's type of transistor, tests must be made on large numbers of transistors. From the results of these tests, minimum and maximum limits must be set on the design parameters so that a statistical minimum of transistors will be outside these limits. This allows the designer economically to choose operating parameters without sacrificing reliability of operation. The sacrifice that must take place is only in allowable number of inputs M and outputs N. The following tests are necessary for NOR circuit design.

Test 1. Back-Collector Current, I_{CBO}

a. Find I_{CBO} at 25 C (degrees centigrade) at a specified collector supply voltage V_{CC}. Set a maximum limit on I_{CBO} for this test.
b. Determine the rise in temperature required to double I_{CBO}. Call this rise W, or
c. Test I_{CBO} at high-temperature limit to determine stability of back current at this temperature. This test is related to test 4

Test 2. Current Gain β, Base to Collector (D-C Beta)

a. A plot of transistor characteristics as obtained by a transistor curve tracer is shown in Fig. 2. The horizontal portion of the collector-current curves are for the saturated "switch-on" condition of operation. On each curve read the highest value of I_c just before the transistor becomes unsaturated. These points as indicated on Fig. 2 give corresponding values of I_c and I_B and the resultant current gain β.
b. Set a minimum limit on β at a specified collector current.

Test 3. Forward-Collector Saturation Voltage Drop, V_{CE}

Again reading from Fig. 2, the voltage drop across the collector and emitter circuit at saturation is obtained from the same indicated points. This saturation voltage will become the maximum "off" signal and should be held to less than 0.25 volt.

Test 4. Noise Figure (NF)

A noise-figure measurement is made at maximum allowable collector voltage. The noise figure is defined as noise above thermal noise at 1,000 cycles with a 1.0-cycle band-pass and is expressed in decibels. The test is made on a standard noise test meter such as the Electronic Research Associates, Inc., Transistor Noise Figure Meter, model *NFT*, and calibrated against a standard noise source. This test is made at a high collector voltage so that the majority of the noise occurs in the collector junction of the transistor. A high noise figure indicates high points on the collector that can cause failure and/or unstable I_{CBO} at both low and high temperatures. This test is therefore a measure of reliability, and transistors with unusually high noise figures should be rejected.

Test 5. Base Input Characteristics, I_B Versus V_{BE}

Using the test circuit of Fig. 3, a curve of I_B versus V_{BE} is plotted. This curve is shown in Fig. 4, for a typical transistor.

Circuit-Design Equations

With the information gained from the foregoing tests, the NOR circuit is designed for reliable operation, allowing adequate margin for drift and aging of the transistor parameters. A maximum number of allowable inputs M and outputs N should be chosen consistent with the margin desired. The outputs N are limited by their loading of the output voltage V_0 down to a point where there is insufficient signal to the following stages. The inputs M are limited chiefly by the input loading, although a crosstalk factor could become important if R_1 were reduced to a very low value.

Fig. 5 shows the circuits considered for both the input-loading and the output-loading cases. The analysis of the input and output circuits of this figure should give definite relations between the transistor parameters, the circuit impedance, and the number of inputs and outputs.

Consider, first, transistor T_1 in Fig. 5. In the cutoff condition, the input signal is zero, or, at least, not more than −0.2 or −0.25 volt since the inputs are the saturated voltages of other conducting transistors. The cutoff collector current is highest at the maximum operating temperature and is given by

$$I_C=\beta(I_{CBO}+I_B)+I_{CBO} \qquad (1)$$

where I_{CBO} is the base-to-collector leakage with the emitter open-circuited, and β is the base-to-collector current gain. If the base current is made negative and equal to the magnitude of I_{CBO}, the first term of the expression is zero, and the collector current is reduced to I_{CBO}. The foregoing expression is not valid for $|-I_B| > I_{CBO}$, and I_{CBO} is the lowest value of I_C attainable. Since the transistor is to remain cut off at the highest operating temperature with zero or just slightly negative input voltage, the necessary I_B must be supplied from the bias source I_T, which can be considered as a current source in most cases.

$$I_T=I_{CBO_{max}}=I_B \text{ and } V_{in}\simeq 0v \qquad (2)$$

The base bias current I_T should be equal to the maximum value of I_{CBO} at the maximum operating temperature for

---

Paper 57-196, recommended by the AIEE Computing Devices Committee and the Industrial Control Committee of the Applications and Industry Division, and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 26, 1956; made available for printing December 6, 1956.

WILLIAM D. ROWE and GEORGE H. ROYER are with the Westinghouse Electric Corporation, in Buffalo, N. Y., and East Pittsburgh, Pa., respectively.

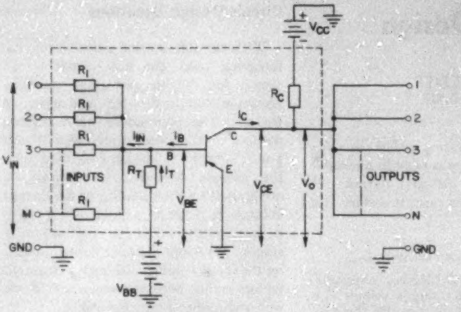


Fig. 1. Transistor NOR circuit

the highest leakage transistor to be used; see test 1.

The maximum marginal loading on the input of a NOR element occurs when a minimum input voltage $V_{0\min}$ is present on one input, and all other inputs are grounded by a saturated transistor as shown in the lower right section of Fig. 5. For this case, I_{in}, the current supplied by the preceding transistor, can be calculated.

$$I_{in}=\frac{V_{BE}-V_{0\min}}{R_1}=SI_{B\min}+\frac{V_{BB}-V_{BE}}{R_T}-\frac{V_{BE}(M-1)}{R_1} \qquad (3)$$

$I_{B\min}$ is determined by the minimum gain of transistor T_2 for the value of I_c to be chosen; see test 2. S is the margin of safety multiplier which allows for future decrease in gain of the transistor, as well as other circuit-parameter tolerances. V_{BE} is a function of base current, but at the low minimum voltage, minimum current case under consideration, $SI_b \leq$ 300 μa (microampere), and V_{BE} is not greater than -0.25 volt; see Fig. 4. Using -0.25 volt as a maximum value of V_{BE}, and knowing also that $|V_{BE}| << |V_{BB}|$ so that $V_{BB}-V_{BE}/R_T \simeq I_T$, the expression for I_{in} reduces to

$$I_{in}=SI_B+I_T+0.25\,\frac{M-1}{R_1}\ \text{at}\ V_{0\min} \qquad (4)$$

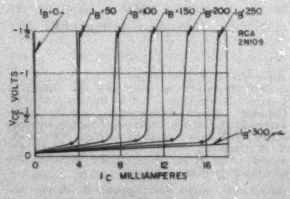


Fig. 2. Common-emitter collector characteristics

The output-loading case, shown in the upper left section of Fig. 5, would be most extreme if all N outputs loaded T_1 similarly to T_2 just calculated. The resultant $V_{0\min}$ would be

$$V_{0\min}=V_{cc}+(NI_{in}+I_{CBO})R_C=-I_{in}R_1+V_{BB} \qquad (5)$$

Solving for N

$$N=\frac{1}{I_{in}}\left(-\frac{V_{cc}-V_{BB}}{R_c}-I_{CBO}\right)-\frac{R_1}{R_c} \qquad (6)$$

The base voltage $V_{BB} < < V_{cc}$ and may be dropped, while R_c is chosen to allow the design maximum value of I_c from the following equation

$$R_cI_c=-V_{cc} \qquad (7)$$

If $I_{CBO} < < I_c$, the expression for N reduces to

$$N=I_c\left(\frac{1}{I_{in}}+\frac{R_1}{V_{cc}}\right) \qquad (8)$$

Substitution of equation 4 in equation 8 gives

$$N=I_c\left(\frac{1}{SI_B+I_T+0.25\,\dfrac{M-1}{R_1}}+\frac{R_1}{V_{cc}}\right) \qquad (9)$$

Equation 9 gives the number of succeeding NOR circuits which may be connected to T_1 under the worst case of loading. It is a function of the following: the transistor parameters I_c, the corresponding I_B, and I_T (or I_{CBO}); the supply voltage V_{cc}; the safety factor S; and two circuit variables, $-R_1$, the input resistor, and M, the number of inputs desired. A high value of V_{cc}, a negative quantity, consistent with the transistor maximum collector voltage, will reduce the negative second term of N. M and S can be stipulated by the designer so that the expression for N becomes a function of only one circuit parameter R_1 and of the transistor to be used. From the results of test 2 and Fig. 2, the corresponding values of I_B and I_c can be obtained, and the safety-factor multiplier S included to give SI_B. Test 1c gives the I_{CBO} maximum design value directly, or it could be calculated from the results of 1a and 1b by use of the following equation

$$_{CBO_{max}}=(II_{co}\ \text{at 25 C})2^{E}$$

$$E=\frac{\text{maximum temperature in C}-25\text{ C}}{W\ \text{in C}} \qquad (10)$$

I_T is designed equal to or somewhat greater than the maximum I_{CBO} encountered with a particular transistor. Fig. 6 shows a family of curves of N versus I_c calculated from equation 9. The different values of R_1 influence the curves, and the maximum value of N occurs at different values of I_c. The final choice of R_1 and I_c involves the compromise of desired values of M and N, the power dissipation in the transistor and the total power dissipation.

The transistor dissipation is a maximum when a lightly loaded transistor has an input from each of it M inputs, and each input supplies only this transistor. For this case

$$I_{B\max}=M\left(\frac{-V_{cc}}{R_c+R_1}\right)$$

$$P_{diss}=P_{base}+P_{coll}=V_BI_{B\max}+V_{CE}I_C \qquad (11)$$

The maximum power-supply dissipation per NOR circuit occurs when the transistor is saturated.

$$P_{supply}=V_{cc}I_c+V_{BB}I_T \qquad (12)$$

The final choice of supply voltages and currents should be guided by keeping all power levels much less than the maximum allowable transistor dissipation.

Crosstalk is another type of loading which can occur between NOR circuits, as shown in Fig. 7. Transistor T_1 is saturated, maintaining a low value of V_{CE}. N-1 of the outputs are connected to other transistors such as T_3 which has the rest of its inputs at full signal. The

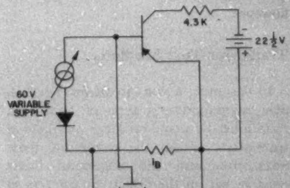


Fig. 3. Test circuit for V_{bIB} curves

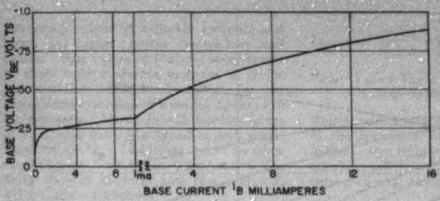


Fig. 4. Base input characteristic of 2N109 as measured by circuit of Fig. 3

base voltage of T_2 is high because of the high base current

$$I_B = (M-1)\left(\frac{-V_{cc}}{R_c+R_1}\right)$$

This high base voltage V_{BE} can cause a current I_e to flow from the collector of T_1 to the base of T_2. A total crosstalk current $I_{CT} = (M\text{-}1)\, I_e$ could flow through the collector of T_1 in addition to its flow to the normal collector current I_c. I_{CT} must be limited to a small percentage of I_c so that T_1 does not become unsaturated and V_{CE} does not rise on the Nth output of T_1, which must still be at zero signal. An upper limit of the value of I_{CT} is obtained by calculating

$$I_B = (M-1)\left(\frac{-V_{cc}}{R_c+R_1}\right),$$

finding V_{BE} from Fig. 4, and using this in the expression

$$I_{CT} = (N-1)\left(\frac{-V_{BE}}{R_1}\right)$$

If $I_{CT} < I_c$ the safety-factor margin in SI_B allows some additional collector current under these extreme crosstalk conditions.

The analysis shown is sufficient to determine an optimum NOR circuit design for any given transistor, subject to the stability or marginal limits imposed by the designer, and the number of logical connections desired. The Appendix shows an example of these calculations for an RCA (Radio Corporation of America) *2N109* transistor.

The frequency of operation of the NOR circuit has not yet been considered. As designed in the Appendix the NOR circuit operates well up to 25 to 40 kc. For many slower control processes, this speed should be sufficient. However, Fig. 8 shows two possible ways of speeding up the response of the circuit. The first is by putting small 100- to 300-micromicrofarad capacitors in parallel with the input resistors. This simple step alone raised the repetition rate of the NOR circuit to 350 to 400 kc. The second speed-up method is described by Baker[1] and uses the subtraction of voltage drops between the silicon and germanium diodes to automatically keep the collector from heavy saturation. This means, of course, that the design values have to be changed to allow for the drop in the silicon diode in the base circuit, and for the higher off-signal value of V_{CE}. This circuit further increases the repetition rate for the *2N109* to 500 or 600 kc at least.

Fifteen hundred NOR elements were constructed and tested More than 1,000 were subjected to an aging cycle of 24 hours at 70 C with supply voltages applied, 24 hours at room temperature with voltage, and 48 hours at 70 C with supply voltage and a square-wave input signal. Only five failures were found in this aging process, and none have failed since in service. Power-supply fluctuations of ±30% and high-ripple percentages have not caused malfunctions.

In summary, this paper presents a method for designing the transistor NOR circuit to include good safety margins for the most severe circuit conditions encountered. This design was applied to a large enough sample to show its reliable operation in actual service.

Appendix. Example of Use of Design Criteria

Test Data, RCA 2N109

Test 1

a. $I_{CBO_{max}} = 5.0\ \mu a$ at $V_{cc} = -25$ volts, 25 C

b. $W = 9$ C

c. $I_{CBO_{max}} \leq 80\ \mu a$ at $V_{cc} = -25$ volts, 65 C

Test 2

Data for I_c versus I_b for lowest gain transistor used is shown in Table I.

Test 3

$V_{CE_{sat}} \leq 0.25$ volt

Test 4

Upper limit on noise figure at $V_{cc} = -25$ volts, and $I_E = 1.0$ ma is 25 decibels above thermal noise.

Test 5

A typical I_B versus V_{BE} curve is shown in Fig. 4.

Choice of Circuit Constants

$$N = I_c\left(\frac{1}{SI_B + I_T + 0.25\,\dfrac{M-1}{R_1}} + \frac{R_1}{V_{cc}}\right) \qquad (9)$$

Evaluation of equation 9 requires: 1. Finding I_T and I_c versus I_B from the test

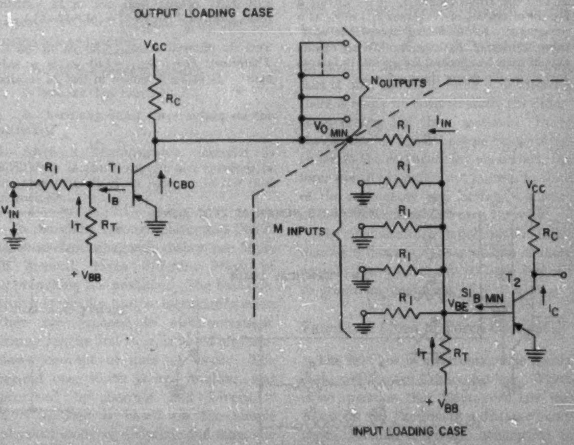


Fig. 5. NOR circuit conditions for calculation of input and output loading

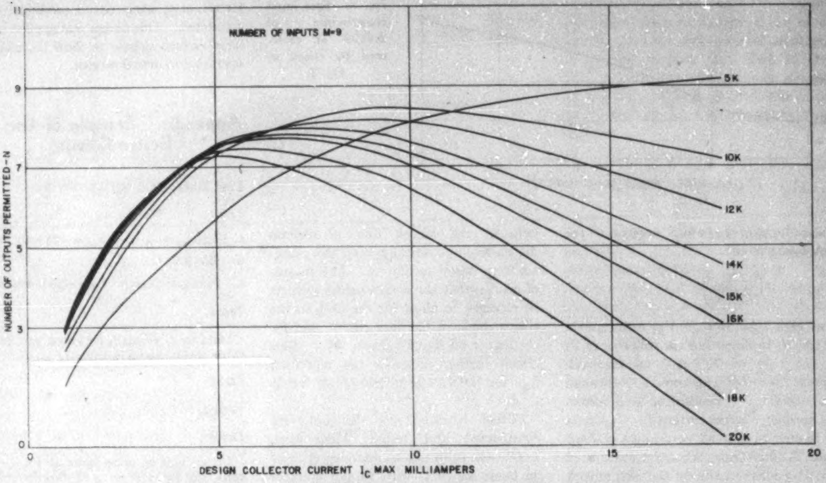


Fig. 6. Graph of number of inputs N versus collector current for 2N109 example

Table I

I_b, μa	20	50	100*	150	200	250	300	400
I_c, ma (milliamperes)	1.02	2.5	5.0*	7.12	9.32	11.4	13.5	17.8
β	51	50	50*	47.5	46.5	45.5	45	44.5

* Minimum d-c β test at $I_c = 5.0$ ma is 50.

data. 2. Selecting V_{cc} for the system. 3. Choosing an acceptable value of inputs M and the safety factor S. 4. Plotting a family of curves of N versus I_c for R_1 as a parameter. 5. Choosing acceptable operating values of R_1 and I_c. These results should then be checked for crosstalk factors and allowable power dissipation.

1. To calculate I_T for 65 C operation use the results of test 1a and 1b in equation 10

$$I_{CBO_{\max}} = (5)\left(2\,\frac{65-25}{9}\right) = (5)2^{4.44} = 110\ \mu a$$

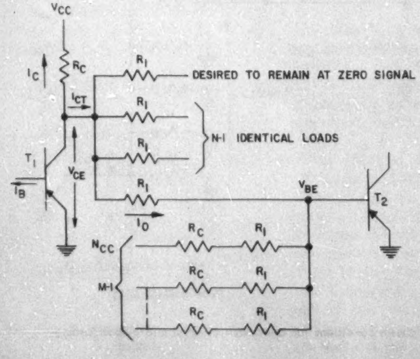
Test 1c showed $I_{CBO_{\max}} \leq 80$ μa at 65 C. Therefore, $I_T = I_{CBO_{\max}} = 110$ μa is a safe figure. I_c vs. I_B is shown in test 2a results.

2. A collector voltage of 20 volts is within the safe operating limits of the transistor for noninductive switching service; $V_{cc} = -20$ volts.

3. Choose $M=9$ as an acceptable number of inputs for each NOR circuit, and use a safety factor $S=2$ for the minimum base current.

4. Fig. 6 shows a family of curves of N versus I_c for $M=9$ and R_1 as a parameter. A high value of N is desirable, consistent with keeping collector and base dissipation low.

5. Fig. 6 shows that the maximum number of outputs N increases as the input resistor R_1 decreases, but this maximum occurs at



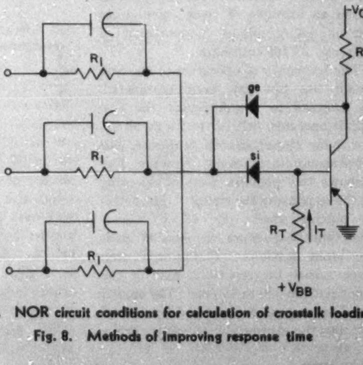


Fig. 7. NOR circuit conditions for calculation of crosstalk loading

Fig. 8. Methods of improving response time

increasing values of collector current, and hence at higher power dissipation and drains. Any values of N above seven seems to show only moderate increase in N for substantial increase in I_c. Thus, for the graph shown, a value of 5 ma for I_c seems to be sufficient. Here, $R_1=15$ kilohms gives $N=7.6$, and there is a broad enough maximum on the $R_1=15$-kilohm curve to allow operation at $N=7$, from $4.2<I_c<10.6$ ma. With $R_1=15$ kilohms and $N=7$, $R_c=20$ volts/5 ma $=4$ kilohms.

Choose $R_c=3{,}900$ ohms, the next lowest Radio-Electronics-Television Manufacturers Association value, so that I_c actually equals 5.13 ma. The bias resistor R_T is calculated from $R_T=V_{BB}/I_T$, and a reasonable choice of V_{BB} is +20 volts, so that it could be supplied from the same transformer windings as V_{cc}.

Thus,

$$R_T=\frac{V_{BB}}{I_T}=\frac{20}{0.11}=180 \text{ kilohms}$$

Checking the power dissipation in the transistor from equation 11 gives

$$I_{B\max}=M\frac{-V_{cc}}{R_c+R_1}=9\frac{(20)}{3.9+15}=9.5 \text{ ma}$$

$$P_{\text{dis}}=V_BI_B+V_{CE}I_C=(0.75)(9.5)+(0.2)(5.13)\simeq 10 \text{ mw}_{\max} \text{ (milliwatts)}$$

This is sufficiently below the 50-mw rating of the transistor to allow for derating at higher operating temperatures. From equation 12 the power supplied to each saturated NOR circuit can be calculated.

$$P_{\text{supply}}=V_{CC}I_C+V_{BB}I_T=(20)(5.13)+(20)(0.110)=105 \text{ mw}$$

Check maximum possible crosstalk current.

$$I_B=(M-1)\frac{-V_{cc}}{R_c+R_1}=\frac{(8)(20)}{3.9+15}=8.5 \text{ ma}$$

$$V_{BE}=-0.75 \text{ volts}$$

$$I_{CT}=(N-1)\frac{-V_{BE}}{R_1}=\frac{(6)(.75)}{15}=300\ \mu\text{a}$$

This amount of increased collector current due to crosstalk is inconsequential.

Summarizing the results of the calculations gives the following:

$R_1=15{,}000$ ohms
$R_c=3{,}900$ ohms
$R_T=180{,}000$ ohms
$V_{cc}=-20$ volts
$V_{bb}=+20$ volts
$I_{c\max}=5.13$ ma
$I_T=0.11$ ma
$M=9$ inputs
$N=7$ outputs

Of course, M or N may still be exceeded in this design if care is taken to assure that all marginal conditions cannot occur simultaneously.

Reference

1. MAXIMUM EFFICIENCY SWITCHING CIRCUITS, R. H. Baker. Presented at the IRE 4.1 subcommittee meeting, Cornell University, Ithaca, N. Y., June 5, 1956. Also *Technical Report no. 110, Lincoln Laboratories*, Massachusetts Institute of Technology, Lexington, Mass.

Forces and Stresses in an Electromagnetic Field

THOMAS H. LEE
MEMBER AIEE

IN MANY engineering problems, it is required to calculate the forces on dielectric and magnetic materials when they are placed in an electromagnetic field. Quite often, it is also desirable to know the force density (the distribution of the forces over the entire volume and the surface) of the material. Methods of calculating the forces and the force densities are given in many books on electromagnetic theory. The treatment of the problem generally leads the reader to believe that the solution to the problem is known and that the equations given are correct. However, if one investigates the equations given in different literatures, one finds that they differ considerably and they give completely different force densities. The author failed to find in the literature a single paper which discusses all the different methods and which points out the most important differences among the theories. The author also found that the differences among the theories are generally not well known to engineers. It is for this reason that this paper is written. Typical questions discussed are as follows:

1. How do the forces originate?
2. Where are they located?
3. Is there any shear force acting on the material?
4. After a ferromagnetic material is saturated, is there any further increase in force as a result of an increase in the field intensity?

In searching through books and literatures on electromagnetic theory, one finds, in general, three different ways of approaching the problem. The first one, which is usually taught to students when they are exposed to electromagnetic theory for the first time, is based on Faraday's concept of lines of force.[1] The second one, which is less familiar, was proposed by Larmor and Livens.[2-4] This method is based on the simple physical model of dielectric and magnetic materials which is sometimes referred to as the Poisson-Kelvin theory of polarization. The third one, which is often reproduced by the authors in Germany and the United States, was suggested by Korteweg, Helmholtz, and others.[5,6] This method can be called the energy method. In the energy method, the variation of the electrostatic or electromagnetic energy of the system due to a displacement is calculated, and this variation of energy, with the sign changed, is taken as the force acting on the system. This variation of energy is in the form of a volume integral and the integrand is taken as the force density, which is quite different from the force density obtained by the Larmor and Livens method.

In the following, the three different methods are discussed and it will be shown that if only the total force is of interest, there is really no need to study the differences among the three methods. From the discussion, it will become evident that as far as the engineers are concerned, the force density is really not known, in spite of the arguments put forward by some authors defending their own theory.[7]

The effect of strain on polarization is not discussed because the problem is more complex, and generally of little interest to practicing engineers.

Faraday's Lines of Force Concept

The first law in electromagnetic theory is the well-known Coulomb's law. There is no question that Coulomb's law was based on the "action at a distance" concept. Michael Faraday, attempting to find the mechanism involved in the transmission of the force, introduced the concept of lines of force, lines pointing along

---

Paper 57-109, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21-25, 1957. Manuscript submitted June 12, 1956; made available for printing November 23, 1956.

THOMAS H. LEE is with the General Electric Company, Philadelphia, Pa.

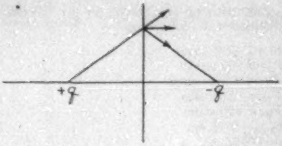


Fig. 1. Two point charges

the direction of the force and acting on charges in space. Faraday visualized the lines of force as having the same properties of elastic bands. He postulated that there is a tension along the line and there is a pressure at right angles to the line. To illustrate the origin of the tension and the pressure, the simple problem or two point charges as shown in Fig. 1 is considered.

The electric intensity E at every point on the mid-plane is at right angles to the plane. The lines of force, therefore, all intersect the mid-plane at right angles. If one cuts a line open at the middle and looks at the left half of the line from the right, the force acting on the line should be toward the right. The force is, therefore, a tension along the line. If the two charges in Fig. 1 are of the same sign, then the electric intensity E at every point on the mid-plane is tangential to the plane. If one looks at one of these lines, one could imagine there is a pressure on the line pushing the lines on the right side of the mid-plane toward the right. It was from this kind of reasoning that Faraday concluded that there was a tension along the lines of force and there was a pressure at right angles to the lines of force. For some time during the development of the theory, it was believed that the Coulomb force had to be transmitted by a certain medium. The all-pervading ether was postulated for this purpose. The force density at every point in the medium should be a function of the magnitude of the charges and the distances from the point to the charges. Therefore, it appeared to be logical to assume that the force density was a function of E. The integral of this function f over any closed surface surrounding one of the two charges must be equal to the total force acting on the charge due to the other one. It can be shown that a possible function for f is,[3] for both tension and pressure,

$$f=\frac{1}{2}\epsilon_0 E^2 \tag{1}$$

For quite some time, it was believed that the medium ether was actually under strain in transmitting the force between charges, just like any elastic medium transmitting mechanical forces. Some authors believed that the phenomena could not be completely understood if the elastic constants of ether were not known. It is now known that the propagation of electrostatic or electromagnetic forces can be realized without any supporting medium; and the nonexistence of ether is completely accepted. However, the concept of tension and pressure acting on the lines of force is, to a degree, still maintained. People are still considering f, as given by equation 1, as tension and pressure which are entities with physical realities. These concepts were also used to calculate the force at the boundary between two dielectric or magnetic materials.

At the boundary between an uncharged dielectric material and free space (Fig. 2), the normal component of the electric displacement and the tangential component of the electric intensity are continuous.

$$D_{n1}=D_{n2}=D_n \tag{2}$$

$$E_{t1}=E_{t2}=E_t \tag{3}$$

According to the hypothesis expressed in equation 1, the tension acting on the left side of the boundary is $1/2\epsilon E^2_{n2}$ and the tension acting on the right side of the boundary is $1/2\epsilon_0 E^2_{n1}$. The resulting tension is

$$T=\frac{1}{2}D_n^2\left(\frac{1}{\epsilon_0}-\frac{1}{\epsilon}\right) \tag{4}$$

Similarly, the resultant pressure is

$$P=\frac{1}{2}E_t^2(\epsilon_0-\epsilon) \tag{5}$$

and the total density of force is

$$T-P=\frac{1}{2}D_n^2\left(\frac{1}{\epsilon_0}-\frac{1}{\epsilon}\right)-\frac{1}{2}E_t^2(\epsilon_0-\epsilon) \tag{6}$$

The foregoing expression is generally considered as the actual force density at the boundary. The surface integral of equation 6 over any closed surface surrounding the dielectric would give the correct total force acting on the dielectric. However, as force density, equation 6 disagrees with that derived by Larmor and Livens.

Larmor and Livens Theory

The method suggested by Larmor and Livens was based on the dipole model of dielectrics. The magnetostatic case and the electrostatic case are exactly alike and it is only necessary to treat one case.

A dipole consists of two equal charges of opposite sign, separated by a very small distance. A dielectric material contains dipoles. When there is no external field acting on the dielectric, these dipoles arrange themselves in random directions so that the average dipole moment per unit volume is zero. An impressed field will, however, orient these dipoles so that there is a resultant dipole moment per unit volume. This resultant dipole moment per unit volume is called the polarization and is generally denoted by P.

DIELECTRIC
(2)
ε
FREE SPACE
(1)
ε₀

Fig. 2. A semi-infinite dielectric

The dipole moment m of a single dipole is defined as a vector quantity, pointing from the negative charge to the positive charge. In a nonuniform electric field, the force acting on the positive charge is different from the force acting on the negative charge, since they are separated by a distance, no matter how small. The components of the force are

$$F_x=m_x\frac{\partial E_x}{\partial x}+m_y\frac{\partial E_x}{\partial y}+m_z\frac{\partial E_x}{\partial z} \tag{7}$$

$$F_y=m_x\frac{\partial E_y}{\partial x}+m_y\frac{\partial E_y}{\partial y}+m_z\frac{\partial E_y}{\partial z} \tag{8}$$

$$F_z=m_x\frac{\partial E_z}{\partial x}+m_y\frac{\partial E_z}{\partial y}+m_z\frac{\partial E_z}{\partial z} \tag{9}$$

In vector notation

$$\mathbf{F}=(\mathbf{m}\cdot\nabla)\mathbf{E} \tag{10}$$

For a dielectric material, the volume force density is, therefore, equal to

$$\mathbf{F}=(\mathbf{P}\cdot\nabla)\mathbf{E} \tag{11}$$

If the dipole model of the dielectric material is correct, then equation 11 describes the physical origin of the force acting on the dielectric. For magnetic materials, the treatment is exactly the same, except P is replaced by M, the magnetization vector or the magnetic moment per unit volume, and E is replaced by H, the magnetic intensity. If the dielectric is charged with a volume density ρ, then the volume force density is

$$\mathbf{F}=(\mathbf{P}\cdot\nabla)\mathbf{E}+\mathbf{E}\rho \tag{12}$$

At the boundary between two dielectrics, equation 12 has to be modified to obtain the surface force density. It has been shown,[3] based on this theory, that if there is no free surface charge at the boundary, the force per unit area at the boundary is

$$\mathbf{F}=\frac{1}{2\epsilon_0}(P_{1n}^2-P_{2n}^2) \tag{13}$$

The total force acting on a dielectric

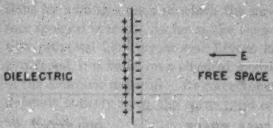


Fig. 3. An electric field normal to the boundary of a semi-infinite dielectric

body imbedded in another dielectric medium is the sum of the surface integral of equation 13 over the bounding surface and the volume integral of equation 12 over the entire volume of the dielectric material.

Equation 13 clearly indicates that the surface force at the boundary is determined by the normal component of polarization. If the field is tangential at the boundary and if the polarization is in the same direction as the electric field, then there is no surface force acting on the boundary.

The Energy Method

The energy method of calculating the forces on dielectric and magnetic materials was proposed by Korteweg and developed by Helmholtz and others. The details of the derivation can be found in several books and will not be repeated here.[8] However, in order to illustrate the major differences among the theories, the essential steps are outlined in the following.

The total mechanical force acting on a dielectric body is found by subjecting every point in the dielectric to an arbitrary displacement and then calculating the change in electrostatic energy due to the displacement. The change in energy due to the arbitrary displacement is composed of three parts: the change in elastic energy, the change in electrostatic energy due to the strain, and the change in electrostatic energy due to the nonhomogeneity of the dielectric. For materials of which the dielectric constants are not affected by strain, only the change in electrostatic energy due to the nonhomogeneity of the dielectric will be important.

The total electrostatic energy in the field is

$$U=\frac{1}{2}\int\epsilon E^2dv \tag{14}$$

and the change in energy due to the displacement is

$$dU_t=\frac{1}{2}\int\delta\epsilon E^2dv \tag{15}$$

If the arbitrary displacement of every point in the dielectric is s, then the following obtains

$$\delta\epsilon=s\cdot\nabla\epsilon \tag{16}$$

and the change in energy is

$$dU_t=\frac{1}{2}\int E^2\nabla\epsilon\cdot s\,dv \tag{17}$$

If there is a distribution of real charges in the dielectric, then the work done by the force acting on the real charges is

$$W=\int\rho\mathbf{E}\cdot s\,dv \tag{18}$$

The total change in energy is

$$\delta U=\int\left(\rho\mathbf{E}-\frac{1}{2}E^2\nabla\epsilon\right)\cdot s\,dv \tag{19}$$

The minus sign in equation 19 indicates that if the electrostatic energy is increased due to the displacement, then the force is opposing such a displacement. If the force density in the dielectric is f, then the total work done by the force is

$$\delta W=\int\mathbf{f}\cdot s\,dv \tag{20}$$

Consequently

$$\int\mathbf{f}dv=\int\left(\rho\mathbf{E}-\frac{1}{2}E^2\nabla\epsilon\right)dv \tag{21}$$

By a well-known law in tensor analysis, the volume integral of equation 21 can be transformed into a surface integral:[9]

$$\mathbf{F}=\int\mathbf{t}\cdot da \tag{22}$$

where

$$\mathbf{t}=\epsilon\mathbf{E}(\mathbf{E}\cdot\mathbf{n})-\frac{\epsilon}{2}E^2 \tag{23}$$

If one concludes from equation 21 that $(\rho\mathbf{E}-1/2E^2\nabla\epsilon)$ is the volume force density, the assumption involved is essentially the recognition of $1/2\epsilon E^2$ as the energy density in an electrostatic field. Although the term energy density is used very frequently, actually, in deriving the total energy in an electrostatic field, there is no mention of the energy density at all. The objection to the idea of energy density has been raised by Mason and Weaver.[10] Stratton[11] has pointed out that it may be questioned whether the term energy density has any physical significance. By the same reasoning, the function $1/2E^2\nabla\epsilon$ has no physical significance, except that its volume integral is equal to the total change in energy and thus is equal to the mechanical force acting on the total dielectric, which is a physically measurable quantity.

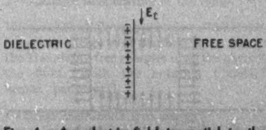


Fig. 4. An electric field tangential to the boundary of a semi-infinite dielectric

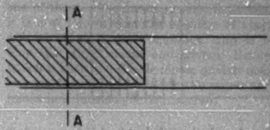


Fig. 5. A long dielectric slab between two conducting plates

Illustrative Examples

To illustrate the difference among the theories, several examples are discussed in this section.

A SEMI-INFINITE DIELECTRIC

Fig. 3 shows a semi-infinite dielectric to the left and free space to the right of the boundary. If a uniform field E, normal to the boundary, is applied, then the dielectric will be polarized as shown. Using equation 6 and since $E_t=0$, the force per unit area is

$$F=\frac{1}{2}D^2\left(\frac{1}{\epsilon_0}-\frac{1}{\epsilon}\right) \tag{24}$$

where

$$D=\epsilon_0E$$

Using the dipole model, one obtains the force density at the boundary from equation 13

$$F=\frac{1}{2\epsilon_0}P_{2n}^2 \tag{25}$$

where

$$P_{2n}=\epsilon_0E\left(1-\frac{\epsilon_0}{\epsilon}\right) \tag{26}$$

Physically, the force can be explained as follows: The force on the negative charges is toward the right and the force on the positive charges is toward the left. The normal field suffers a discontinuity in crossing the boundary and consequently, the force acting on the negative charges is different from the force acting on the positive charges. There is, therefore, a resultant force acting on the boundary. The electric intensity at the right side of the boundary is higher than that at the left side and the resultant force acting on the dielectric is, therefore, toward the right.

Equations 24 and 25 do not give the same force density except for $\epsilon=\infty$. If

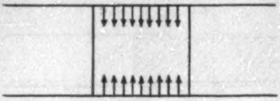


Fig. 6. A short dielectric slab between two conducting plates

one uses equation 23, one obtains the following

$$t=\frac{1}{2}\epsilon_0E^2 \qquad (27)$$

This is different from either equation 24 or equation 25.

Consider now the same boundary except that a uniform tangential field is applied as shown in Fig. 4. In this case, the answers are

1. By equation 6

$$P=\frac{1}{2}E_t^2(\epsilon_0-\epsilon)$$

2. By equation 13

$$F=0$$

3. By equation 23

$$\mathbf{t}=-\frac{1}{2}\epsilon_0E_t^2$$

Using the dipole model, it can be seen from Fig. 4 that there is no force at the boundary.

A Dielectric Slab Between Two Parallel Plates

Fig. 5 shows a dielectric slab inserted into a parallel-plate capacitor. The capacitor plates are large so that the field at the center is tangential to the boundary of the dielectric. At the boundary, the normal component of polarization is, therefore, equal to zero. According to Larmor and Livens' theory the surface tension at the boundary is zero. There is, however, a resultant body force acting on the dielectric to move it toward the right. This force is a body force and has a density of

$$(\mathbf{P}\cdot\nabla)\mathbf{E}$$

distributed over the whole dielectric. This force is nearly zero at the center of the capacitor and increases to a maximum at a location near the left end of the capacitor. One thus may say that the force is originated in a region near the left end of the capacitor plates. At the horizontal boundary between the dielectric and capacitor plates, there is a discontinuity of electric intensity and hence there is a pressure acting on the dielectric material. If this dielectric is a fluid, then the pressure will be transmitted to the boundary between the dielectric and the empty space; but if the dielectric is a solid body, then this pressure is not completely transmitted. If now one applies equation 6, one obtains a surface force density of

$$\frac{1}{2}E^2(\epsilon-\epsilon_0)$$

at the boundary between the dielectric and free space.

If one clamps down the dielectric at *AA*, then one theory indicates that the left portion is under compression and the right portion is not stressed. The other theory indicates that the right portion is under tension, but the left portion is not stressed.

If the dielectric is much smaller than the capacitor plates and is sitting in the middle of the capacitor as shown in Fig. 6, there is no resultant force on the dielectric.

There is a pressure acting on the horizontal boundaries of the dielectric as shown by the Larmor and Livens' theory. Again, if this dielectric is a fluid, this pressure will be transmitted to the two vertical boundaries, but if the dielectric is a rigid solid, the pressure is not transmitted. According to equation 6, the dielectric is stressed as shown in Fig. 7. The disagreement is, therefore, obvious.

A Semi-infinite Dielectric Slab and a Point Charge

As the last example, the case of a semi-infinite dielectric slab and a point charge *q* as shown in Fig. 8 will be considered.

This is a well-known problem. The solution by the method of images[11] gives the total force acting on the dielectric as

$$F=\frac{q^2(\epsilon_0-\epsilon)}{16\pi\epsilon_0a^2(\epsilon+\epsilon_0)} \qquad (28)$$

The problem can also be solved by the three different methods.

Using Lines of Force Concept

The tension per unit area at the boundary is

$$T=\frac{\epsilon^2q^2a^2}{8\pi^2(\epsilon+\epsilon_0)^2(a^2+r^2)}\left(\frac{1}{\epsilon_0}-\frac{1}{\epsilon}\right) \qquad (29)$$

and the pressure is as given in the following equation

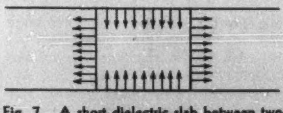


Fig. 7. A short dielectric slab between two conducting plates

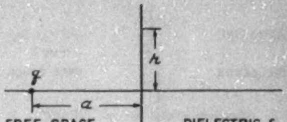


Fig. 8. A point charge and a semi-infinite dielectric

$$P=\frac{q^2r^2(\epsilon-\epsilon_0)}{8\pi^2(\epsilon+\epsilon^2)_0(a^2+r^2)^3} \qquad (30)$$

When one integrates the function $T-P$ over the entire boundary of the dielectric, one obtains equation 28.

Using Larmor and Livens' Theory

According to Larmor and Livens' theory, the force per unit area at the boundary is

$$f=\frac{q^2a^2(\epsilon-\epsilon_0)^2}{8\pi^2\epsilon_0(\epsilon+\epsilon_0)^2(a^2+r^2)^3} \qquad (31)$$

and the volume force density is

$$f_v=\frac{1}{2}(\epsilon-\epsilon_0)\frac{\partial E^2}{\partial x} \qquad (32)$$

When one integrates equation 31 over the entire boundary and equation 32 over the entire volume of the dielectric and adds them, the result is the same as equation 28. Equation 31 is, however, different from $T-P$ given by equations 29 and 30. If one sets $\epsilon=\infty$, then equation 31 becomes identical to equation 29 and equation 30 becomes zero. Thus, for a dielectric material with an infinite dielectric constant, the two theories again lead to identical results.

Using Energy Method

If one uses equation 23, then the total force is given by

$$F=\int\mathbf{t}\cdot da \qquad (33)$$

where

$$\mathbf{t}=\epsilon_0E_x^2-\frac{\epsilon_0E^2}{2}$$

$$=\frac{q^2a^2\epsilon^2}{4\pi^2\epsilon_0^2(\epsilon+\epsilon_0)^2(a^2+r^2)^3}-\frac{q^2(r^2\epsilon_0^2+a^2\epsilon^2)}{8\pi^2\epsilon_0(\epsilon+\epsilon_0)^2(a^2+r^2)^3} \qquad (34)$$

The total force obtained is again the same as equation 28.

It thus becomes obvious that the functions as given by equations 6 and 23 have significance only when integrated over the entire bounding surface. This integral gives the total force acting on the dielectric. One can easily change these functions by adding a term of which the surface integral is zero. As far as the equation proposed by Larmor and Livens is concerned, it is based on a physical model of the dielectric material. Its correctness depends entirely upon the correctness of the model.

Conclusions

Although many physical theories deal with the behavior of elementary particles, experimental observations can only be made on ensembles. While the theory may treat the magnetic field of a charged particle, it is only possible to measure the magnetic field of a complete loop of current. Thus, for a single physical phenomenon, there may be several theories. Each one of the theories may differ from the other theories in the differential form; but when integrated to predict the performance of the whole system, all theories may converge to the same result. Many different concepts have been proposed for electromagnetic theory, but the only available experiments are those of Coulomb, Ampere, and Faraday. It can be seen that the problem discussed in this paper is one of the many cases in the development of physical theory.

Several useful conclusions can be drawn for the magnetic case from the foregoing:

1. For unsaturated ferromagnetic materials, the permeability is so much higher than that of air that it is a good approximation that the force is entirely a surface traction. This is similar to the force acting on a perfect conductor.

2. The surface force is always normal to the surface and there is no shear force acting on the surface.

3. The conventional force formula $B^2A/2\mu$ implicitly assumed that: (*a*) B is so high that the difference between it and M is negligible; and (*b*) μ is so high that the difference between it and μ_0 of air is the same as μ.

4. In calculating the resultant force on an electromagnet, the value $1/2\,\mu B^2A$ should be evaluated over all surfaces of the armature or the plunger, not over one surface alone.

5. The permeability μ in the foregoing formula is equal to B/H, not the differential permeability dB/dH.

6. After the armature or plunger of an electromagnet is completely saturated and then if the excitation in the coil is further increased, there can be a continuous increase in the total force acting on the armature or the plunger if the field caused by the current in the coil is nonuniform in the region where the armature or plunger is located. This is not easily seen from the energy method but is clearly indicated by the Larmor and Livens' theory. This force is simply the volume integral $\int(\mathbf{M}\cdot\nabla)\mathbf{H}dV$. The magnetization M is now a constant. If a uniform field is superposed, there is no increase in force.

7. The three theories discussed differ in results if the material is not absolutely rigid. Unfortunately, there is no experimental information to indicate the correctness of any one of them. For ferromagnetic materials, it is now known that the problem of magnetostriction is quite complex, and it is true that the Larmor and Livens' theory which is based on simple dipole models does not yield correct results for the stress and strain in a ferromagnetic material.

References

1. Static and Dynamic Electricity (book), W. R. Smythe. McGraw-Hill Book Company, Inc., New York, N. Y., 1950, p. 15.

2. Critique of von Helmholtz's Theory of Electric Stresses, J. Larmor. *Philosophical Transactions*, Royal Society, London, England, 1897, p. 280.

3. Theory of Electricity (book), G. H. Livens. Cambridge University Press, London, England, 1918, p. 177.

4. On the Mechanical Relations of Dielectric and Magnetic Polarization, G. H. Livens. *Philosophical Magazine*, London, England, July-Dec. 1916, p. 162.

5. Encyklopaedie der Mathematischen Wissenschaften, F. Pockels. Leipzig, Germany, vol. V, pt. II, 1906, pp. 350–92.

6. Electromagnetic Theory (book), J. A. Stratton. McGraw-Hill Book Company, Inc., chap. II, 1941.

7. On the Mechanical Forces in Dielectrics, W. B. Smith-White. *Philosophical Magazine*, London, England, 1948, p. 466.

8. Smythe. *op. cit.*, p. 16.

9. Stratton. *op. cit.*, p. 100.

10. The Electromagnetic Field (book), M. Mason, W. Weaver. University of Chicago Press, Chicago, Ill., 1929, p. 266.

11. Stratton. *op. cit.*, p. 110.

12. Electromagnetism (book), J. C. Slater, N. H. Frank. McGraw-Hill Book Company, Inc., 1947, p. 47.

◆

Discussion

Roland M. Lichtenstein (Rensselaer Polytechnic Institute, Troy, N. Y.): We should like to supplement the author's presentation with an exposition of the thermodynamic method for calculating stresses.

This method will work only when restricted to materials that are, at least locally, in thermodynamic equilibrium. This restriction precludes the discussion of such interesting problems as the evaluation of stresses in the presence of electromagnetic waves. However, even with this restriction, the method can handle many problems of practical importance.

In order to hold the exposition simple we shall consider the example of fluids consisting of a single constituent (not a mixture). The fluid is assumed to be an insulator and charge-free. The state of no charge within the liquid is expressed by

$$\operatorname{div} \mathbf{D} = 0 \tag{35}$$

where $\mathbf{D}$ is the electric displacement.

A concrete embodiment of the situation we should like to consider is a completely enclosed room with metallic walls. Within the room there are distributed N metallic conductors carrying the charges Q_i ($i=1, 2, \ldots N$). The space in between the conductors is filled with fluid. The outside wall is assumed to carry the charge $-(Q_1+Q_2+\ldots Q_N)$, so that there is no field beyond the outside wall. We shall assign the order number 0 to the conductor that constitutes the outside wall. Then

$$Q_0 = -(Q_1+Q_2+\ldots Q_N)$$

The whole arrangement is assumed to be at some given constant absolute temperature T. This isothermal condition may be achieved if the room is immersed in a heat bath of large heat capacity and if all changes, such as motions of the conductors, to be contemplated later proceed slowly enough so that at any time the temperature can equalize by heat conduction.

We consider the Helmholtz free energy F of the fluid defined by

$$F = U - TS \tag{36}$$

where

U = energy of the fluid
S = entropy of the fluid
T = absolute temperature of the fluid

We make the following postulate: The Helmholtz free energy F of the fluid is the sum of the Helmholtz free energies of all the parts of the fluid. With the notation ("molar" means "per mole")

f = molar Helmholtz free energy
v = molar volume

the postulate is expressed by

$$F = \iiint_{\text{fluid}} f\frac{1}{v}\,d\tau \tag{37}$$

Here $d\tau$ is the volume element. One arrives at equation 37 by realizing that a volume element $d\tau$ contains $1/v d\tau$ moles.

The postulate states essentially that the free energy may be localized. Logically, there is no compelling reason for doing this. However, the decision to employ localization is in accord with the spirit of thermodynamics. If we did not use the postulate, we would deprive ourselves of the possibility of applying thermodynamics. The decision to use the postulate is, in the end, based on the practical usefulness of the theory.

Because of the symmetry inherent in the structure of fluids, the molar Helmholtz free energy f can depend only on T, v, and the absolute value $|\mathbf{D}|$ of the electric displacement $\mathbf{D}$ (or on $\mathbf{D}\cdot\mathbf{D}/2 = D^2/2$). We ignore the dependence on T, since we agreed to hold T constant. We thus have

$$f = f\left(v, \frac{\mathbf{D}\cdot\mathbf{D}}{2}\right) = f\left(v, \frac{D_x^2+D_y^2+D_z^2}{2}\right) \tag{38}$$

Here D_x, D_y, D_z are the cartesian components of $\mathbf{D}$.

Let us assume that the charges Q_i on the conductors and the total number of moles n of the fluid are given. We have the foling two equations

$$\iint\limits_{\text{conductor } i} \mathbf{D}\cdot d\mathbf{a}=Q_i \text{ for } i=0, 1, \ldots N \tag{39}$$

and

$$n=\iiint\limits_{\text{fluid}} \frac{1}{v}\, d\tau \tag{40}$$

The integral in equation 39 is taken over the surface of each conductor. The vectorial surface element, pointing into the liquid, is denoted by $d\mathbf{a}$.

Our first task is to determine the molar volume v and the electric displacement **D** as a function of location. We do this by starting with the requirement that the total Helmholtz free energy F must be a minimum as we vary v and **D** by small quantities δv and $\delta\mathbf{D}$ while we hold fixed:

1. The shapes and locations of the conductors.
2. The charges on the conductors.
3. The amount of fluid.
4. The state of no charge within the fluid.

This requirement leads first of all to the following variational problem for the four field functions v, D_x, D_y, D_z.

$$\left.\begin{aligned} &\iiint f\left(v, \frac{D_x^2+D_y^2+D_z^2}{2}\right)\frac{1}{v}d\tau \\ &\qquad\qquad\qquad = \text{minimum} \\ &\text{with the restricting conditions} \\ &\iiint \frac{1}{v} d\tau = n = \text{const} \\ &\text{div } \mathbf{D}=\frac{\partial D_x}{\partial x}+\frac{\partial D_y}{\partial y}+\frac{\partial D_z}{\partial z}=0 \end{aligned}\right\} \tag{41}$$

For a moment let us denote the four field functions by φ_α, $(\alpha=1, 2, 3, 4)$. The variational problem yields the following four Euler-Lagrange equations

$$[h]_{\varphi_\alpha}=0 \text{ for } \alpha=1, 2, 3, 4 \tag{42}$$

Here

$$h=\frac{f}{v}-g\frac{1}{v}-\psi\left(\frac{\partial D_x}{\partial x}+\frac{\partial D_y}{\partial y}+\frac{\partial D_z}{\partial z}\right) \tag{43}$$

where g is a Lagrange multiplier that is independent of location, and where ψ is a Lagrange multiplier that does depend on location $[h]_{\varphi_\alpha}$ is the variational derivative defined by

$$\begin{aligned} [h]_{\varphi_\alpha}=\frac{\partial h}{\partial \varphi_\alpha}-\frac{\partial}{\partial x}\left(\frac{\partial h}{\partial\left(\frac{\partial \varphi_\alpha}{\partial x}\right)}\right)- \\ \frac{\partial}{\partial y}\left(\frac{\partial h}{\partial\left(\frac{\partial \varphi_\alpha}{\partial y}\right)}\right)-\frac{\partial}{\partial z}\left(\frac{\partial h}{\partial\left(\frac{\partial \varphi_\alpha}{\partial z}\right)}\right) \end{aligned} \tag{44}$$

Equations 42 then become

$$\frac{1}{v}\frac{\partial f}{\partial v}-\frac{1}{v^2}f+\frac{1}{v^2}g=0$$

$$\frac{1}{v}\frac{\partial f}{\partial(D^2/2)}D_x+\frac{\partial \psi}{\partial x}=0$$

$$\frac{1}{v}\frac{\partial f}{\partial(D^2/2)}D_y+\frac{\partial \psi}{\partial y}=0$$

$$\frac{1}{v}\frac{\partial f}{\partial(D^2/2)}D_z+\frac{\partial \psi}{\partial z}=0$$

or

$$f-v\frac{\partial f}{\partial v}=g=\text{constant throughout the fluid} \tag{45}$$

$$\frac{\mathbf{D}}{v}\frac{\partial f}{\partial(D^2/2)}=-\text{grad } \psi \tag{46}$$

Equation 45 expresses the important result that quantity

$$g=f-v\frac{\partial f}{\partial v}$$

is a constant throughout the liquid. This quantity is called the molar Gibbs free energy.

Equation 46 states that the vector

$$\mathbf{E}=\frac{\mathbf{D}}{v}\frac{\partial f}{\partial(D^2/2)} \tag{47}$$

is the negative gradient of some potential ψ.

$$\mathbf{E}=-\text{grad } \psi \tag{48}$$

The Euler-Lagrange equations arise if the permissible variations of those field functions whose space derivatives appear in h (that is D_x, D_y, D_z) are such that these variations vanish at the boundaries of the fluid. Actually, of course, we were not required to hold D_x, D_y, D_z fixed at the boundaries. The only condition is that the charges on the conductors are fixed. This is condition 2, which is expressed by equation 39. Thus we are permitted a larger selection of variations than we so far have used. Hence we should expect more severe restrictions on the potential function ψ. In order to find those additional restrictions we start with the variation δF of the Helmholtz free energy F.

$$\begin{aligned} \delta F=\iiint\limits_{\text{fluid}}\Bigg[\left(\frac{1}{v}\frac{\partial f}{\partial v}-\frac{1}{v^2}f\right)\delta v+\frac{1}{v}\times \\ \frac{\partial f}{\partial(D^2/2)}(D_x\delta D_x+D_y\delta D_y+D_z\delta D_z)\Bigg]d\tau \end{aligned}$$

With the use of equations 45–48 this expression becomes

$$\begin{aligned} \delta F&=\iiint\limits_{\text{fluid}}\left[g\left(-\frac{1}{v^2}\right)\delta v+\mathbf{E}\cdot\delta\mathbf{D}\right]d\tau \\ &=g\delta n-\iiint(\text{grad } \psi)\cdot\delta\mathbf{D}d\tau \end{aligned}$$

But δn must vanish. Then, with a simple vector relation,

$$\delta F=-\iiint[\text{div }(\psi\delta\mathbf{D})d\tau-\psi \text{ div } \delta\mathbf{D}]d\tau$$

But div $\delta\mathbf{D}$ must vanish. We convert the surviving integral into surface integrals by means of Gauss' theorem. Then (remember that the vectorial surface element d**a** points into the liquid),

$$\delta F=+\sum_{i=0}^{N}\iint\limits_{\text{conductor } i}\psi\delta\mathbf{D}\cdot d(\mathbf{a}) \tag{49}$$

Let us define an average value of ψ over conductor i by

$$<\psi>_i=\frac{\iint\psi|d\mathbf{a}|}{\iint|d\mathbf{a}|} \tag{50}$$

The integrals are taken over conductor i. We may then write equation 49 in the form

$$\begin{aligned} \delta F=\sum_{i=0}^{N}\Bigg[<\psi>_i\iint\delta\mathbf{D}\cdot d\mathbf{a}+ \\ \iint(\psi-<\psi>_i)\delta\mathbf{D}\cdot d\mathbf{a}\Bigg] \end{aligned} \tag{51}$$

Suppose that we make the choice

$$\delta\mathbf{D}\cdot d\mathbf{a}=-(\psi-<\psi>_i)|d\mathbf{a}|\epsilon_i \tag{52}$$

for each conductor surface i. Here the ϵ_i are arbitrary small positive numbers. The choice, equation 52, fulfills the condition that the charge on each conductor remains unchanged since, with equation 52,

$$\delta Q_i=\iint\limits_{\text{conductor } i}\delta\mathbf{D}\cdot d\mathbf{a}=0 \tag{53}$$

By inserting equations 52 and 53 into equation 51 we get

$$\delta F=-\sum_{i=0}^{N}\epsilon_i\iint\limits_{\text{conductor } i}(\psi-<\psi>_i)^2,|d\mathbf{a}|, \tag{54}$$

This expression is negative, contrary to the requirement that F be a minimum, unless for each conductor $\psi=<\psi>$, that is, unless over each conductor surface the potential ψ is a constant.

Up to now we have concluded that $\mathbf{E}=-$grad ψ and that ψ is a constant over each conductor surface. A consequence is that **E** is normal to the surface. Without loss of generality we may choose $\psi=0$ on the outside wall.

Now let us vary the charges on the inner conductors by δQ_i, $i=1, \ldots.$N. The work input for this variation is

$$\delta W=\sum_{i=1}^{N}V_i\delta Q_i \tag{55}$$

where V_i is the conductor voltage with respect to the outside wall. This work input must show up as an increase δF of the Helmholtz free energy.

$$\delta W=\delta F \tag{56}$$

If we calculate δF as before, we arrive at equation 49. Since we already concluded that ψ is constant over each conductor and since we set $\psi=0$ on the outside wall, equation 49 may be written in the form

$$\delta F=\sum_{i=1}^{N}\psi_i\delta Q_i \tag{57}$$

When we combine equation 55 to equation 57, and when we realize that the δQ_i are arbitrary, we conclude that at each conductor

$$\psi_i=V_i \tag{58}$$

Hence not only is **E** the negative gradient of some potential, but this potential is also equal to the conductor voltage at each surface. The vector **E** then has all the properties of an electric field. We shall therefore identify the vector **E** with the electric field.

To this point then we have shown that:

1. The molar Gibbs free energy $g \equiv f - v\partial f/\partial v$ is constant throughout the fluid.
2. The electric field **E** is given by $\mathbf{E} \equiv \mathbf{D}/v\, \partial f/\partial(D^2/2)$.
3. The electric field has a potential $\mathbf{E} = -\text{grad}\,\psi$.
4. On each conductor ψ is constant and equal to the conductor voltage. The electric field is normal to the conductor.

We yet have to finish the task of finding the four field functions v, D_x, D_y, D_z. This task may be reduced to the calculation of two field functions ψ and p, where p is defined by

$$p = -\frac{\partial f}{\partial v} + \mathbf{D}\cdot\mathbf{E} \tag{59}$$

First we must show how v and **D** may be determined from p and ψ. With equation 59 we have

$$g = f - v\frac{\partial f}{\partial v} = f + pv - v\mathbf{D}\cdot\mathbf{E} \tag{60}$$

and

$$dg = df + p\,dv + v\,dp - \mathbf{D}\cdot\mathbf{E}\,dv + v\mathbf{D}\cdot d\mathbf{E} + v\mathbf{E}\cdot d\mathbf{D} \tag{61}$$

But

$$df = \frac{\partial f}{\partial v}dv + \frac{\partial f}{\partial(D^2/2)}\mathbf{D}\cdot d\mathbf{D}$$

or

$$df = -p\,dv + \mathbf{D}\cdot\mathbf{E}\,dv + v\mathbf{E}\cdot d\mathbf{D} \tag{62}$$

On combining equations 61 and 62 we get

$$dg = v\,dp - v\mathbf{D}\cdot d\mathbf{E} \tag{63}$$

Let us express the molar Gibbs free energy g as a function of p and $E^2/2$. (For symmetry reasons g can depend only on p and the absolute value of **E** or $\mathbf{E}^2/2$. and, of course the temperature T.)

From equation 63 there follows

$$v = \frac{\partial g}{\partial p} \tag{64}$$

$$\mathbf{D} = -\frac{\mathbf{E}}{v}\frac{\partial g}{\partial(E^2/2)} \tag{65}$$

Equations 64 and 65 tell us how to determine v and **D**, once we know p and ψ, (remember that $\mathbf{E} = -\text{grad}\,\psi$). The two field functions p and ψ are determined from the equations.

$$g = \text{const} \tag{66}$$

$$\text{div}\,\mathbf{D} = -\text{div}\left(\frac{\mathbf{E}}{v}\frac{\partial g}{\partial(E^2/2)}\right) = 0 \quad \text{where } \mathbf{E} = -\text{grad}\,\psi \tag{67}$$

The value of the constant in equation 66 depends on the number of moles n. The boundary values of ψ are the conductor voltages, which are given either directly or in terms of the conductor charges.

We finally come to the determination of the stress tensor $\bar{S}$. Let us shrink the inner conductors and expand the outside wall by small displacements. The displacement of the general surface point is denoted by $\delta\mathbf{r}$.

We hold fixed the following:

1. The state of no charge within the fluid.
2. The charges on the conductors.
3. The amount of fluid.

In what follows, all volume integrals are taken over the old volume of the fluid and all surface integrals are taken over all of the old surfaces. We assume that the conductors are made of foils that can be stretched and flexed without effort. The work input for the indicated variation is then

$$\delta W = -\iint \delta\mathbf{r}\cdot\bar{S}\cdot d\mathbf{a} \tag{68}$$

The minus sign in equation 68 comes from the convention that the vectorial surface element $d\mathbf{a}$ points into the fluid. The work input must show up as an increase δF of the Helmholtz free energy F. We have (remember that this time also the boundaries are varied),

$$\delta F = \iiint\left[\left(\frac{1}{v}\frac{\partial f}{\partial v} - \frac{1}{v^2}f\right)\delta v + \frac{1}{v}\frac{\partial f}{\partial(D^2/2)}\mathbf{D}\cdot\delta\mathbf{D}\right]d\tau - \iint\frac{1}{v}f\,\delta\mathbf{r}\cdot d\mathbf{a}$$

or

$$\delta F = g\iiint\left(-\frac{1}{v^2}\right)\delta v\,d\tau + \iiint\mathbf{E}\cdot\delta\mathbf{D}\,d\tau - \iint\frac{1}{v}f\delta\mathbf{r}\cdot d\mathbf{a} \tag{69}$$

Furthermore

$$\delta n = \iiint\left(-\frac{1}{v^2}\right)\delta v\,d\tau - \iint\frac{1}{v}\,\delta\mathbf{r}\cdot d\mathbf{a} \tag{70}$$

But, according to condition 3,

$$\delta n = 0 \tag{71}$$

On combining equation 69 to equation 71 we obtain

$$\delta F = \iint\frac{g-f}{v}\,\delta\mathbf{r}\cdot d\mathbf{a} + \iiint\mathbf{E}\cdot\delta\mathbf{D}\,d\tau$$

or

$$\delta F = -\iint\frac{\partial f}{\partial v}\delta\mathbf{r}\cdot d\mathbf{a} + \iiint\mathbf{E}\cdot\delta\mathbf{D}\,d\tau \tag{72}$$

But

$$\begin{aligned}\iiint\mathbf{E}\cdot\delta\mathbf{D}\,d\tau &= -\iiint\delta\mathbf{D}\cdot\text{grad}\,\psi\,d\tau\\ &= -\iiint\text{div}\,(\psi\delta\mathbf{D})\,d\tau + \iiint\psi\,\text{div}\,\delta\mathbf{D}\,d\tau\end{aligned}$$

According to condition 1, we have $\text{div}\,\delta\mathbf{D} = 0$. After converting the surviving integral on the right-hand side into a surface integral, we get

$$\iiint\mathbf{E}\cdot\delta\mathbf{D} = +\iint\psi\delta\mathbf{D}\cdot d\mathbf{a} = \sum_{i=0}^{N}\psi_i\iint\delta\mathbf{D}\cdot d\mathbf{a} = \sum_{i=0}^{N}\psi_i\delta Q_i$$

where δQ_i is the change of charge on conductor i. According to condition 2, we have $\delta Q_i = 0$. Then

$$\iiint\mathbf{E}\cdot\delta\mathbf{D}\,d\tau = 0 \tag{73}$$

and equation 72 becomes

$$\delta F = -\iint\frac{\partial f}{\partial v}\delta\mathbf{r}\cdot d\mathbf{a} \tag{74}$$

When we combine equations 68 and 74 (remember $\delta F = \delta W$) and realize that the displacements $\delta\mathbf{r}$ are arbitrary, we obtain

$$\bar{S}\cdot d\mathbf{a} = \frac{\partial f}{\partial v}d\mathbf{a} \tag{75}$$

We now make the postulate that the stress tensor $\bar{S}$ can depend only on v and **D**. Furthermore, since the fluid is in equilibrium, the force density must vanish, which gives

$$\nabla\cdot\bar{S} = 0 \tag{76}$$

Let us see whether the following expression $\bar{I}$ is the idemfactor

$$\bar{S} = \left(\frac{\partial f}{\partial v} - \mathbf{D}\cdot\mathbf{E}\right)\bar{I} + \mathbf{DE} \tag{77}$$

or, with equation 59,

$$\bar{S} = -p\bar{I} + \mathbf{DE} \tag{78}$$

fulfils equation 72 and the boundary condition, equation 71. (The easiest way for obtaining equation 77 is through the analysis of a capacitor consisting of two concentric metallic spheres between which there is fluid. In order to conserve space we shall not reproduce this analysis here.) We have

$$\begin{aligned}\bar{S}\cdot d\mathbf{a} &= \frac{\partial f}{\partial v}d\mathbf{a} + \mathbf{DE}\cdot d\mathbf{a} - \mathbf{D}\cdot\mathbf{E}\,d\mathbf{a}\\ &= \frac{\partial f}{\partial v}d\mathbf{a} + \mathbf{E}\times\mathbf{D}\times d\mathbf{a}\end{aligned}$$

Now **D** is parallel to **E**, and on the surfaces **E** is parallel to $d\mathbf{a}$. Hence $\mathbf{D}\times d\mathbf{a}$ vanishes, and we have

$$\bar{S}\cdot d\mathbf{a} = \frac{\partial f}{\partial v}d\mathbf{a}$$

which checks equation 75. From equation 78 we get

$$\nabla\cdot\bar{S} = -\text{grad}\,p + \mathbf{E}\,\text{div}\,\mathbf{D} + \mathbf{D}\cdot\nabla\mathbf{E}$$

But div **D** vanishes, with equation 65 we get

$$\nabla\cdot\bar{S} = -\text{grad}\,p - \frac{\partial g}{\partial(E^2/2)}\frac{1}{v}\mathbf{E}\cdot\nabla\mathbf{E}$$

and with equation 64,

$$\nabla\cdot\bar{S} = -\frac{1}{v}\left(\frac{\partial g}{\partial p}\text{grad}\,p + \frac{\partial g}{\partial(E^2/2)}\mathbf{E}\cdot\nabla\mathbf{E}\right) \tag{79}$$

Now

$$\text{grad}\,E^2/2 = \mathbf{E}\times\text{curl}\,\mathbf{E} + \mathbf{E}\cdot\nabla\mathbf{E}$$

But $\text{curl}\,\mathbf{E} = -\text{curl grad}\,\psi$ vanishes. Then equation 79 becomes

$$\nabla\cdot\bar{S} = -\frac{1}{v}\left(\frac{\partial g}{\partial p}\text{grad}\,p + \frac{\partial g}{\partial(E^2/2)}\text{grad}\,E^2/2\right)$$

or

$$\nabla\cdot\bar{S}=-\frac{1}{v}\operatorname{grad} g \tag{80}$$

Since g is constant throughout the fluid, grad g vanishes, so that

$$\nabla\cdot\bar{S}=0$$

which checks equation 76.

Equation 78 shows that, in the absence of an electric field, the quantity p is the pressure. When an electric field is present, p has no physical meaning other than that given by the definition, equation 59.

Let us consider an example. We expand g $(p, E^2/2($ in a Taylor serie swith respect to $E^2/2$ carrying only the first two terms.

$$g=g_0(p)+g_1(p)\frac{E^2}{2} \tag{81}$$

With equations 64 and 65 we obtain

$$v=\frac{dg_0}{dp}+\frac{E^2}{2}\frac{dg_1}{dp} \tag{82}$$

$$\mathbf{D}=-\frac{\mathbf{E}}{v}g_1 \tag{83}$$

In the particular case of an incompressible fluid v must be a constant, say v_0. Then

$$g_0=pv_0 \tag{84}$$

except for an additive constant, which we may set equal to zero without loss of generality. Furthermore, g_1 must be independent of p. Then equation 83 becomes

$$\mathbf{D}=-\frac{g_1}{v_0}\mathbf{E} \tag{85}$$

It is customary to denote the constant $-g_1/v_0$ by ϵ, where ϵ is the permittivity. Then

$$\mathbf{D}=\epsilon\mathbf{E} \tag{86}$$

and

$$g=pv_0-v_0\epsilon\frac{E^2}{2}$$

The stress tensor becomes

$$\bar{S}=-\frac{g_0}{v_0}\bar{I}+\epsilon\left(\mathbf{EE}-\frac{E^2}{2}\bar{I}\right) \tag{87}$$

We note that the first term in equation 87 is constant throughout the fluid. The second term resembles the stress tensor for vacuum; the vacuum permittivity ϵ_0 is replaced by the permittivity ϵ of the incompressible fluid.

When we have to deal with a compressible fluid, it is better to expand the molar Helmholtz free energy $f(v, D^2/2)$ in a Taylor series with respect to $D^2/2$. Carrying only the first two terms, we get

$$f=f_0(v)+f_1(v)\frac{D^2}{2} \tag{88}$$

It is customary to write $f_1(v)$ in the form $v/\epsilon(v)$. Then equation 88 becomes

$$f=f_0(v)+\frac{v}{\epsilon}\frac{D^2}{2} \tag{89}$$

Equation 47 gives the following

$$\mathbf{E}=\frac{\mathbf{D}}{\epsilon(v)} \tag{90}$$

The quantity $\epsilon(v)$ is the permittivity. From equation 77 we get

$$\bar{S}=\left[\frac{\partial f_0}{\partial v}+\frac{D^2}{2}\left(\frac{1}{\epsilon}-\frac{v}{\epsilon^2}\frac{\partial\epsilon}{\partial v}\right)-\mathbf{D}\cdot\mathbf{E}\right]\bar{I}+\mathbf{DE}$$

or

$$\bar{S}=\left(\frac{\partial f_0}{\partial v}-\frac{\epsilon E^2}{2}\frac{\partial\log\epsilon}{\partial\log v}\right)\bar{I}+\epsilon\left(\mathbf{EE}-\frac{E^2}{2}\bar{I}\right) \tag{91}$$

A readily observable affect is the dependence of the molar volume v on location. We have, with sufficient accuracy,

$$v=v_{D=0}+\frac{D^2}{2}\left(\frac{\partial v}{\partial\frac{D^2}{2}}\right)_g \tag{92}$$

The partial derivative in equation 92 is to be taken at constant g, since we already know that g is constant throughout the liquid. With equation 89 we have

$$g=f-v\frac{\partial f}{\partial v}=f_0+\frac{v}{\epsilon}\frac{D^2}{2}-v\times\frac{df_0}{dv}-v\frac{D^2}{2}\left(\frac{1}{\epsilon}-\frac{v}{\epsilon^2}\frac{\partial\epsilon}{\partial v}\right)$$

or

$$g=f_0-v\frac{df_0}{dv}+\frac{v}{\epsilon}\frac{D^2}{2}\frac{\partial\log\epsilon}{\partial\log v} \tag{93}$$

We take the total differential of equation 93 and set $D^2/2$ equal to zero (or consider $D^2/2$ small). Then

$$dg=-v\frac{d^2f_0}{dv^2}dv+\frac{v}{\epsilon}\frac{\partial\log\epsilon}{\partial\log v}d(D^2/2)$$

Thus

$$\left(\frac{\partial v}{\partial\frac{D^2}{2}}\right)_g=\frac{1}{\epsilon\frac{d^2f_0}{dv^2}}\frac{\partial\log\epsilon}{\partial\log v} \tag{94}$$

The isothermal compressibility x in the absence of a field is defined by

$$x=-\frac{1}{v}\frac{\partial v}{\partial p}=-\frac{1}{v}\left(\frac{\partial p}{\partial v}\right)^{-1}=\frac{1}{v}\left(\frac{d^2f_0}{dv^2}\right) \tag{95}$$

Then equation 94 becomes

$$\left(\frac{\partial v}{\partial\frac{D^2}{2}}\right)_g=\frac{vx}{\epsilon}\frac{\partial\log\epsilon}{\partial\log v} \tag{96}$$

If ϵ increases with decreasing v, then v decreases with increasing field. On inserting equation 94 into equation 92 we get

$$v=v_{D=0}+\frac{vx}{\epsilon}\frac{D^2}{2}\frac{\partial\log\epsilon}{\partial\log v} \tag{97}$$

or

$$v=v_{D=0}+vx\epsilon\frac{E^2}{2}\frac{\partial\log\epsilon}{\partial\log v} \tag{98}$$

The change of molar volume in an electric field (this is the electrostriction effect) is directly observable. However, the stress tensor is not directly observable, except by some very artificial means such as the introduction of cracks. Yet the stress tensor is a valuable concept. For it allows to predict accelerations in dynamic problems if we make the assumption that in a dynamic situation the stress tensor depends on v and **D** in the same way as in a static situation. This assumption is within the spirit of Newton's second law and is usually, though not always, justified. When dissipative effects—such as viscosity—are important, the assumption is obviously not correct. Yet the stress tensor is still a valuable concept. For by experimentation we can find out how the static stress tensor must be corrected so that the dynamic situation may be handled.

In conclusion we want to state once more the postulates that enabled us to evaluate the stress tensor for fluids:

Postulate 1. The free energy may be localized.

Postulate 2. The stress tensor depends only on the molar volume, the electric displacement **D** and the temperature.

From the observable stress at the surface we could infer the stress tensor by way of the equation

$$\nabla\cdot\bar{S}=0$$

An intermediate result that was of great importance for the further development of the theory was that the molar Gibbs free is constant throughout the fluid. This result is well known in thermodynamics.

T. H. Lee: Dr. Lichtenstein's discussion is a very interesting one. It illustrates another significant application of this very powerful tool—the method of thermodynamics. I often feel that the thermodynamics course taught in engineering colleges is not broad enough to give the students a good picture of what thermodynamics can do. Electrical engineers frequently refuse to study thermodynamics because their impression is that this is an unrelated field for them. However when they face such problems as conduction of electricity in gases, then they discover that thermodynamics is not something which is useful only to mechanical engineers and physicists. I am very happy that Dr. Lichtenstein took the effort to present this additional method of calculating stresses.

I wonder whether it is necessary to even postulate the localization of the free energy. It seems to me that one can start with equation 37 and proceed from there by assuming, for physical reasons, that f must be a function of v and **D** so that its volume integral is equal to the Helmholtz free energy of the system. One would arrive at the same results without giving f too much physical significance. The stress tensor obtained by this means obviously has the same significance as the stress tensor discussed in the paper.

Moment Detection and Coding

J. J. SLADE, JR.
NONMEMBER AIEE

L. F. NANNI
NONMEMBER AIEE

S. FICH
MEMBER AIEE

D. A. MOLONY
NONMEMBER AIEE

Synopsis: This paper describes a method of detecting the characteristics of an electrical pulse by means of moments that are integrals of the instantaneous pulse amplitude and integral powers of time. Application of this method to detection in band-limited systems in the presence of noise is considered. Theoretical calculations are presented in the form of curves showing the required signal-to-noise ratios for various speeds of signalling. These curves illustrate the inherent flexibility of moment detection, and make possible a comparison between moment and amplitude detection.

Definition of Moments

THE APPLICATION of moments in classical mechanics to determine the center of gravity, the moment of inertia, and other properties of mechanical systems is well known. These computations involve the product of integral powers of the distance measured from some reference point and area, volume, mass, etc. In such applications the system is specified, and the moments are calculated to determine its response to external excitation. However, if the system is sufficiently restricted or the number of moments is sufficiently large, its nature can be determined to any desired degree of accuracy from a knowledge of the moments. A restricted mechanical configuration consisting of a given mass at a variable distance from a fixed reference point is completely specified by the first moment which is a linear function of the distance. More complex configurations require a greater number of moments for specification. The use of moments to specify a function has a mathematical basis. A time-limited function can be represented by a Gram-Charlier series that is somewhat analogous to a Fourier series, and in which the coefficients are moments of the given function.[1]

Moments can also be used to determine a configuration of electrical signals consisting of one or more d-c pulses restricted to exist at specified periods of time measured from a reference of zero time. This is the model to be considered in the analysis that is to follow. Such a configuration could be made to follow a code and convey information. When the configuration consists of a single pulse of constant amplitude and width existing at various prescribed values of time, it is the electrical analog of the mass at variable distance. As in the mechanical case, this electrical configuration is uniquely determined by the first moment. The moment can be used to detect the code, and this type of detecting will be termed "moment detection."

Inasmuch as the moments used for detection involve integral powers of time, they may be called "temporal moments." The temporal moment of order k, of a function of time, $f(t)$, calculated in the interval from $t=0$ to $t=T$ is defined as follows:

$$m_k(T,f)=\int_0^T t^k f(t)dt \qquad (1)$$

When $k=0$,

$$m_0(T,f)=\int_0^T f(t)dt \qquad (2)$$

where $m_0(T,f)$ is the area or content under the curve of $f(t)$ from $t=0$ to $t=T$. The first three moments of order, 0, 1, and 2 of a rectangular pulse are shown in Fig. 1. These moments can be computed by means of conventional electronic integrators.[2] Referring to equation 2, it is noted that a zero-order moment detector is simply an integrator.

Application of Moments to Detection

From the point of view of information theory, a pulse may be considered as an elementary signal that carries a minimum set of information parameters, such as peak amplitude, relative position, area, spread, etc. The pulse need not be rectangular. In the case of time-limited pulses, all of these parameters can be determined by moments as defined by equation 1. A moment detector of zero order can detect information transmitted by a pulse of constant position and variable height or width. In this case only the area is variable. If pulses of equal height and width, but variable position are used, the zero-order moment detector fails when the period of integration extends over several allowed pulse positions, and the first moment is required for detection. This situation is illustrated in Fig. 2. Moments are well suited for the detection of pulse code modulation. In Fig. 3 are shown the codes of four of the possible 32 levels of a conventional 5-digit binary system consisting of five spaces to be left vacant or filled by rectangular pulses. All four levels have the same zero moments. The first moment can be used to detect levels A and B, but since levels C and D have the same first moment, the second moment is required to detect these levels. The first three moments are sufficient to detect all 32 levels. Combinations of zero-order moments taken over different time intervals have been used to detect these levels.[3] Before the moments are calculated, the effects of pulse contamination by noise and wave-form degradation due to bandwidth restrictions will be discussed.

Effect of Noise

When a signal is impressed upon the input of a practical communication system, the output consists of two parts. One part, $v(t)$, represents the input signal distorted by the system in a predictable manner, and the other, $n(t)$, is random noise. Assuming that the noise is directly additive, the output can be expressed as

$$f(t)=v(t)+n(t) \qquad (3)$$

The moment recorded by a moment detector at the output is given by

$$\begin{aligned} m_k(T,f)&=\int_0^T t^k[v(t)+n(t)]dt \\ &=m_k(T,v)+m_k(T,n) \qquad (4) \end{aligned}$$

Because of the randomness of the noise $m_k(T,f)$ is a random variable, and the moments are now only statistically predictable. The various values of $m_k(T,f)$ will tend to group around a central value designated as $E[m_k(T,f)]$ where E represents the expected or average value of $m_k(T,f)$. The spread or dispersion about this central value is measured by the standard deviation or standard error of estimate, designated by σ. [illegible] a normal distribution, the mean or central value $\pm\sigma$ includes 68%

Paper 56-963, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956, and re-presented at the Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted June 20, 1956; made available for printing February 6, 1957.

J. J. SLADE, JR., L. F. NANNI, S. FICH, and D. A. MOLONY are with Rutgers University College of Engineering, New Brunswick, N. J.

The theory of moment detection described in this paper represents a portion of the work conducted at Rutgers University under a research project entitled "Theoretical and Experimental Research in Communication Theory and Application," sponsored by the Signal Corps, Department of the Army, since June 15, 1951 under Contract Nos. *DA-36-039-sc-15314*, *42703*, and *64704*. This sponsorship is gratefully acknowledged. Moments were mentioned in the first progress report dated August 25, 1951, and expanded in subsequent reports.

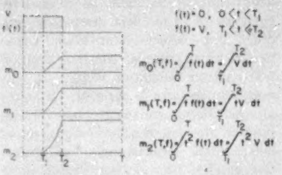

V
f(t)
m_0
m_1
m_2
T_1
T_2
T
$f(t)=0,\ 0<t<T_1$
$f(t)=V,\ T_1<t\leqslant T_2$
$m_0(T,f)=\int_0^T f(t)\,dt=\int_{T_1}^{T_2} V\,dt$
$m_1(T,f)=\int_0^T t\,f(t)\,dt=\int_{T_1}^{T_2} tV\,dt$
$m_2(T,f)=\int_0^T t^2 f(t)\,dt=\int_{T_1}^{T_2} t^2 V\,dt$

Fig. 1. First three moments of a rectangular pulse

of all possible values of the function, and the mean $\pm 2\sigma$ includes 95% of all values. The calculation of the standard deviation is of fundamental importance in the evaluation of moment detection in the presence of noise.

The variance or square of the standard deviation is given by

$$\sigma^2(y)=E(y^2)-[E(y)]^2 \tag{5}$$

Substituting $y=m_k(T,f)$ in equation 5, the variance of the moment of the signal plus the noise is given by

$$\sigma^2[m_k(T,f)]=E[m_k^2(T,f)]-\{E[m_k(T,f)]\}^2 \tag{6}$$

Since $v(t)$ is not a random signal

$$E[m_k(T,f)]=m_k(T,v) \tag{7}$$

It follows from equation 4 that

$$E[m_k^2(T,f)]=m_k^2(T,v)+2m_k(T,v)E[m_k(T,n)]+E[m_k^2(T,n)] \tag{8}$$

If it is assumed that the average noise amplitude is zero

$$E[m_k(T,n)]=\int_0^T t^k E[n(t)]dt=0 \tag{9}$$

Substitution of equations 7 to 9 into equation 6 gives

$$\sigma^2[m_k(T,f)]=\sigma^2[m_k(T,n)] \tag{10}$$

In other words, the variance of the kth moment of the output signal is equal to the variance of the kth moment of the noise.

It follows from equations 5 and 10 that

$$\begin{aligned}\sigma^2[m_k(T,f)]&=E[m_k^2(T,n)]\\&=E[\int_0^T t^k n(t)dt]^2\\&=\int_0^T\int_0^T t^k u^k E[n(t)\times(u)]dtdu\end{aligned} \tag{11}$$

After considerable algebraic manipulation, it is shown that when T is sufficiently large, equation 11 reduces to[4]

$$\sigma^2[m_k(T,f)]=\frac{B_0T^{2k+1}}{2k+1} \tag{12}$$

where B_0 is twice the content of the autocorrelation function, and is given by

$$B_0=2\int_0^\infty \varphi_n(\tau)d\tau \tag{13}$$

and $\varphi_n(\tau)$ is the autocorrelation function of the noise.

Using the Wiener-Khintchine relations discussed by Rice,[5] B_0 in equation 13 can be expressed in terms of the average noise power and the bandwidth of the communication system. These relations state that the autocorrelation function and the power spectrum form a Fourier transform cosine pair. In other words

$$\varphi_n(\tau)=\int_0^\infty \Phi_n(f)\cos 2\pi f\tau df \tag{14}$$

where $\Phi_n(f)$ is the power spectrum of the noise. If the noise is assumed to be "white," i.e., if the power spectrum is assumed to be uniform

$$\Phi_n(f)=\Phi_n(0)=\text{a constant} \tag{15}$$

When the bandwidth of the system is W, equation 14 becomes

$$\varphi_n(\tau)=\Phi_n(0)\int_0^W \cos 2\pi f\tau df=\frac{\Phi_n(0)}{2\pi\tau}\sin 2\pi W \tag{16}$$

In a system of bandwidth W,

$$N^2=W\Phi_n(0)$$

$$\Phi_n(0)=\frac{N^2}{W} \tag{17}$$

where N^2 is the average noise power and N is the rms noise amplitude. Substitution of equations 16 and 17 into equation 13 gives the following result:

$$B_0=\frac{2N^2}{W}\int_0^\infty \frac{\sin 2\pi\tau W}{2\pi\tau}d\tau=\frac{N^2}{2W} \tag{18}$$

When equation 18 is substituted into equation 12, the following expression for the variance of the moment of the output signal is obtained:

$$\sigma_k=\sigma[m_k(T\ f)]=\frac{NT^{k+\frac{1}{2}}}{\sqrt{2(2k+1)W}} \tag{19}$$

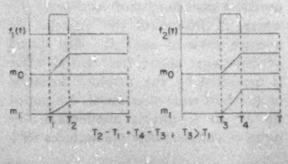


Fig. 2 (left). Moments of pulses of constant area and variable position

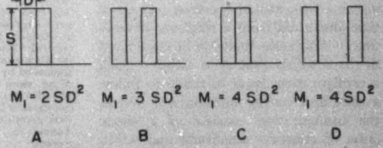


Fig. 3 (right). Four possible levels of a 32-level code

Equation 19 shows that when N and W are constant σ_k is proportional to $T^{k+1/2}$. In general, the moment of order k is proportional to T^{k+1}. It follows that the coefficient of variation, or the ratio of the standard deviation to the moment in question, is inversely proportional to the square root of the period of integration. Thus, the variance caused by noise is reduced when the period of integration is increased.

Effect of Bandwidth Restriction

When the bandwidth of a communication system is limited, the shape of the original pulse, considered as a function of time, is always modified. When a rectangular pulse is passed through an ideal low-pass filter, a finite build-up time is required for the output to reach its maximum value and, instead of falling abruptly to zero, the output approaches zero in a series of oscillations of small amplitude. The net effect of filtering is to increase the effective width of the pulse and to introduce oscillations of small amplitude, frequently referred to as "tails." The moments of these tails would be superposed with some probability distribution upon the moment of the pulse. Because of these moments the detector would give an indication when no pulse is present, and would indicate a moment different from the true moment when a pulse exists. This effect combines with the noise interference to distort the detected moments.

The undesirable effects of the tails can be virtually eliminated, as shown in Fig. 4, which represents the output of an ideal filter of bandwidth W when two very narrow pulses having an amplitude ratio of 0.6 are applied with a time separation of $1/2W$. Note how the tails of the second pulse almost completely cancel those from the first pulse. If the width of the input pulse is k/W where $k<1$, the output of the filter is given by

$$f(t)=\frac{\sin 2\pi W\left(t-\frac{k}{2W}\right)}{n\left(t-\frac{k}{2W}\right)}+$$

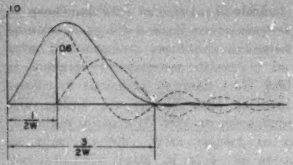


Fig. 4. Almost tailless output from ideal filter

$$0.6\,\frac{\sin 2\pi W\left(t-\frac{k+1}{2W}\right)}{\pi\left(t-\frac{k+1}{2W}\right)} \qquad (20)$$

where $f(t)$ is shown in Fig. 4. The effective width of the pulse is $3/2W$.

Detection of Single Pulse in Presence of Noise

In any system of coding that is to be detected by means of moments, there exists a minimum increment, Δm_k, that must be detected through noise. It is assumed that an error is committed whenever the kth moment of the noise exceeds $\Delta m_k/2$. Using a representation of noise as given by Rice, it is shown that the kth moment of noise in a band-limited system with a flat spectrum has a Gaussian distribution with a mean value of zero and a standard deviation given by equation 19.[6,7] The probability, α_k, of a noise moment exceeding $\Delta m_k/2$, is then given by

$$\alpha_k=\frac{1}{\sigma_k\sqrt{2\pi}}\int_{\Delta m_k/2}^{\infty} e^{-\frac{x^2}{2\sigma_k^2}}\,dx$$
$$=\frac{1}{\sqrt{2\pi}}\int_{x_{\alpha_k}}^{\infty} e^{-\frac{x^2}{2}} \qquad (21)$$

where

$$x_{\alpha_k}=\frac{\Delta m_k}{2\sigma_k} \qquad (22)$$

When a value of α_k is assumed in equation 21, x_{α_k} can be found from probability tables. Substitution of σ_k as given by equation 19 into equation 22 gives equation 23 as follows:

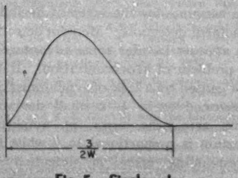


Fig. 5. Single pulse

$$x_{\alpha_k}=\frac{\Delta m_k\sqrt{2(2k+1)W}}{2NT^{k+\frac{1}{2}}} \qquad (23)$$

Inspection of equation 21 shows that α_k, the probability of error, increases as x_{α_k} decreases. It follows from equation 23 that for given values of W and T, x_{α_k} decreases when Δ_{m_k} decreases or k increases. The moment increments become smaller when the pulses are crowded to increase the speed of signalling, and higher-order moments are required to detect the levels in certain multilevel codes. These factors are brought out in the calculations that follow.

Consider the detection of the single tailless pulse shown in Fig. 5 by means of zero-order moments when the period of integration is equal to the pulse width. If S_0 is the average value of the pulse during the period of integration, the zero-order moment is given by

$$m_0=\frac{3}{2W}S_0=TS_0 \qquad (24)$$

Substitution of $\Delta m_0/2=3S_0/4W$, $T=3/2W$, and $k=0$ in equation 23 yields

$$x_{\alpha_0}=\frac{3S_0}{4WN}\,\frac{\sqrt{2W}}{\sqrt{3/2W}}=\frac{S_0}{2N}\sqrt{3} \qquad (25)$$

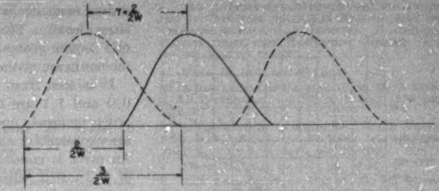
Equation 25 is expressed in decibels as

$$20\log x_{\alpha_0}=10\log\frac{S_0}{N}-1.2494 \qquad (26)$$

It is found that when 20 log S_0/N equals 12.5 db (decibels), $x_{\alpha_0}=3.652$ and $\alpha_0=0.00012$. It follows that $2W/3$ pulses per second can be detected with a probability of error equal to approximately 10^{-4} when the ratio of the average signal of the rms noise is 12.5 db.

This means that if a pulse train of $2W/3$ pulses per second contaminated by white noise is passed through a filter of bandwidth W, the signal-to-noise ratio at the filter input must be such that the signal to noise ratio at the output is at least 12.5 db if the probability of error in zero-order moment detection is 10^{-4}. Table I shows the results of calculations made for speeds of signalling ranging from $2W/3$ to $W/50$ pulses per second. Table I shows that when the speed of signalling is reduced, the required signal-to-noise ratio is also reduced. In amplitude detection the required signal-to-noise ratio is a function only of the probability of error, and when $\alpha=10^{-4}$, the ratio of peak signal to rms noise is 17.4 db. This comparison illustrates the inherent flexibility of moment detection.

Fig. 6. Crowded pulses



To attain higher speeds of signalling it is necessary to crowd the pulses. In other words, a second pulse must begin before the preceding one ends. Fig. 6 shows a new pulse starting $1/2W$ seconds before the end of the preceding one. The distance between successive peaks is reduced to $T=2/2W$, and this is the new period of integration. The effect of crowding, in addition to increasing the speed of signalling, is to decrease the moment quanta and increase the required signal-to-noise-ratio for a given probability of error. Pulse crowding, as shown in Fig. 6, can be increased until the sum of the moments of the preceding and succeeding pulses equals the moment of the given pulse in the reduced period of integration. This reduces the moment quantum to zero and occurs in the neighborhood of $2TW=0.87$. Table II gives the speeds of signalling and signal-to-noise ratio in the case of pulse crowding.

Detection of Multiple-Pulse Groups

To achieve higher speeds of signalling than can be attained by using crowded single pulses, it is necessary to use multiple-pulse groups consisting of two or more pulses existing during the period

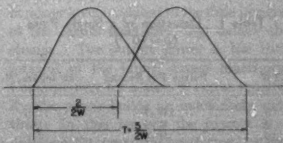


Fig. 7. Two-pulse group

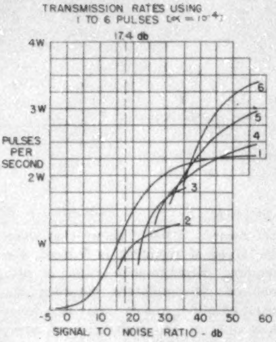


Fig. 8. Transmission rates using 1 to 6 pulses

of integration. To detect all of the combinations that may arise with multiple-pulse groups, moments of higher order than m_0 are required. In Fig. 7 are shown two pulses existing in a period $T=5/2W$. The zero-order moment of level 1,0 corresponding to the presence of the first pulse and the absence of the second is

$$m_0(0,1)=S_0\int_0^{3T/5} dt=\frac{3}{5}S_0T \qquad (27)$$

The zero-order moment of level 0,1 corresponding to the absence of the first pulse and the presence of the second pulse is also $3S_0T/5$. It follows that the zero-order moment is insufficient to detect these levels. The first-order moment of level 1,0 can be expressed as

$$m_1(1,0)=S_1\int_0^{3T/5} t\,dt=\frac{9}{50}S_1T^2 \qquad (28)$$

where

$$S_1=\frac{\int_0^{3T/5} tf(t)dt}{\int_0^{3T/5} tdt} \qquad (29)$$

and $f(t)$ is the equation of the pulse. For the wave form used S_1 is nearly equal to S_0. It can be shown in a similar manner that the first-order moment for level 0,1 is $21S_1T^2/50$. Hence, the first-order moment is sufficient to detect these levels. The four possible levels of a 2-pulse group and the corresponding moments are given in Table III.

It is seen from Table III that levels 0,0 and 1,1 are detected by the zero-order moment, and levels 1,0 and 0,1 are detected by the first-order moment. An error is committed in the detection of level 0,0 when the noise moment exceeds

$$\frac{\Delta m_0}{2}=\frac{3}{10}S_0T \qquad (30)$$

The probability of committing this error will be denoted by α_0. Level 1, 1 will be in error if the moment of the noise is less than $-\Delta m_0/2$. Because of symmetry, the probability of this error is also α_0. The effect of noise in both m_0 and m_1 must be considered in calculating the probability of error in level 1,0. The probability of error in level 1,1 due to m_0 is $2\alpha_0$ because an error is committed if the moment of the noise exceeds $\Delta m_0/2$ or is less than $-\Delta m_0/2$. An error is committed in m_1 if the first moment of the noise exceeds

$$\frac{\Delta m_1}{2}=\frac{1}{2}\left(\frac{21}{50}-\frac{9}{50}\right)S_1T^2=0.12\ S_1T^2 \qquad (31)$$

The probability of this error is denoted by α_1. The combined probability of error in level 1,0 is

$$\alpha_{1,0}=2\alpha_0+\alpha_1-2\alpha_0\alpha_1 \qquad (32)$$

The probability of error in level 0,1 is the same as for level 1,0. Assuming that the four levels occur with equal probability, the over-all probability of error is given by ·

$$\alpha=\frac{1}{4}\alpha_0+\frac{1}{4}(\alpha_0+\alpha_1-2\alpha,\ \alpha_1)+\frac{1}{4}(2\alpha_0+\alpha_1-2\alpha_0\alpha_1)+\frac{1}{4}\alpha_0 \qquad (33)$$

$$=1.5\alpha_0+0.5\alpha_1-\alpha_0\alpha_1$$

(In applying moment detection to an actual code, the level having the highest probability of error can be made to represent the code having the lowest probability of occurrence, etc.)

Following the method used for a single pulse, it can be shown that when 20 $\log_{10}S_1/N=18$, $\alpha_1=0.00012$, $\alpha_0=10^{-6}$ and $\alpha=10^{-4}$. Since two pulses exist in a period $T=5/2W$, it follows that $4W/5$ pulses per second can be detected with a probability of error equal to 10^{-4} when the signal to noise ratio is 18 db. Similar calculations for pulse groups consisting of 3, 4, 5, and 6 pulses have been made and are graphically shown in Fig. 8, which also includes the data for single pulses. The number associated with each curve indicates the number of pulses in the group.

Table III. Levels of a 2-Pulse Group

Level	m_0	m_1
0,0	0	
1,0	$3S_0T/5$	$9S_1T^2/50$
0,1	$3S_0T/5$	$21S_1T^2/50$
1,1	$6S_0T/5$	

Comparison of Amplitude and Moment Detection

The capabilities and limitations of moment detection are illustrated by the curves shown in Fig. 8. All the curves have been plotted for a probability of error equal to 10^{-4}. For this rate of error, the ratio of peak signal to rms noise in amplitude detection is 17.4 db. When the rate of signalling is less than about $1.2W$, the required ratio of average signal to rms noise in moment detection is less than 17.4 db, as shown by curve 1 which is drawn for a single pulse. As the speed of signalling is decreased, this curve extends into the region of negative decibel signal-to-noise ratio. For a speed of signalling of $W/50$ pulses per second, the signal to noise ratio is −3 db. Fig. 8 also shows that speeds in excess of $2W$ pulses per second are attainable provided that sufficient signal-to-noise ratio is available. The higher speeds require the use of both first- and second-order moments. In amplitude detection the required signal to noise ratio is a function only of the probability of error; in moment detection the signal-to-noise ratio is a function both of the probability of error and the speed of signalling. This accounts for the inherent flexibility of moment detection. Moment detection also appears to offer some advantage in the problem of synchronization. If the error caused by a lack of synchronization is assumed to be δ, a small deviation about zero, the error in amplitude detection is proportional to δ while the corresponding error in detection by zero-order moments or integration is

Table I. Detection of a Single Pulse ($\alpha=10^{-4}$)

2TW	Pulses Per Second	20 $\log_{10}S_0/N$
3	$2W/3$	12.5
4	$W/2$	11.5
10	$W/5$	7.5
20	$W/10$	4.5
100	$W/50$	−3.0

Table II. Detection of a Single Pulse with Crowding ($\alpha=10^{-4}$)

2TW	Pulses Per Second	20 $\log_{10}S_0/N$
2	W	15.0
3/2	$4W/3$	18.5
5/4	$8W/5$	21.5
1	$2W$	29.5
7/8	$16W/7$	55.0

proportional to δ^2. As far as the signal-to-noise ratio for a fixed probability of error is concerned, amplitude detection is superior to moment detection for speeds of signalling between $1.2W$ and $2W$ pulses per second, and moment detection appears to offer an advantage at other speeds. With respect to a signal-to-noise ratio of 17.4 db at a speed of $2W$ pulses per second, it appears that moment detection holds most promise of improvement in cases of low signal-to-noise ratio where low speeds of signalling are tolerable and in cases of high signal-to-noise ratio where high speeds of signalling are required.

References

1. Mathematical Methods of Statistics (book), H. Cramer. Princeton University Press, Princeton, N. J., 1946, p. 223.

2. Theoretical and Experimental Research in Communication Theory and Application, Final Report, J. J. Slade, Jr., S. Fich, D. A. Molony, L. F. Nanni. *Contract no. DA-36-039-sc-15314*, Mar. 18, 1954, pp. 52–56.

3. The Integration Method of Detection and Its Application to the Transmission of Teletype Signals, H. F. Harmuth. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 10, 1954, pp. 286–92.

4. Slade, Jr., et al, *loc. cit.*, p. 44.

5. A First Look at Random Noise, S. O. Rice. *AIEE Transactions*, vol. 75, pt. I, May 1956, pp. 128–31.

6. Mathematical Analysis of Random Noise, S. O. Rice. *Bell System Technical Journal*, New York, N. Y., vol. 23, 1944, pp. 282–332.

7. Slade, Jr., et al, *loc. cit.*, pp. 48–49.

8. The Philosophy of P.C.M., B. M. Oliver, J. R. Pierce, C. E. Shannon. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 36, 1948, pp. 1324–31.

Dynamics and Control of Nuclear Power Plants

A. B. VAN RENNES
NONMEMBER AIEE

J. C. SIMONS, JR.
NONMEMBER AIEE

T. S. GRAY
FELLOW AIEE

THE EVOLUTION of the control philosophy and consequent design of the control system for a nuclear central-station power plant reflects a merger of present experience in two areas; namely, control of existing hydrocarbon-fueled central-station plants, and control of mobile, production, and research type nuclear reactors.

The extrapolations to the present problem are greatest in the nuclear area. In comparison with hydrocarbon-fueled central-station systems there may be some increase in capacity, an entirely different source of heat for the boilers, and a considerable variation in the fueling procedure with fissionable fuel. Previous relevant experience with nuclear reactors is derived from small mobile steam-generating power plants, from relatively high-power production reactors, and from high-flux research reactors. Nearly all these installations are externally controlled and are fueled only while shut down. In contrast, many of the presently conceived central-station reactors are self-regulating and are designed so that refueling operations may be carried on while the reactor is continuously delivering power.

This paper presents some of the control philosophy that dictates the manner in which control is to be accomplished, and some of the system design considerations that are common to many reactor plants incorporating heat exchangers.

Methods of Control in Fossil- and Fissile-Fueled Plants

The combination of a nuclear reactor and a turbogenerator in an assembly for the purpose of generating central-station electric power poses some unique problems in control. The problems arise chiefly from the fact that a nuclear reactor is unlike a fossil-fueled plant, in which the fuel residence time is at most a few minutes.

In conventional plants, the heat source is made to deliver heat in accordance with the demand for steam power through adjustment of the rate of feed of fuel to the burners. An increase in power demand results in a slightly lowered steam pressure, which through an automatic combustion controller increases the fuel rate sufficiently to return the steam pressure to its normal value. Because of the short fuel residence time, the heat rate closely follows the fuel rate.

On the other hand, once a nuclear reactor is fueled initially, it may not need refueling for a period of months or years. Hence control of a nuclear reactor through control of the fissionable fuel input is impractical. Enough fuel to last a long time must be installed all at once in order that the chain reaction take place at all.

In many nuclear plant concepts, the reactor is made to deliver heat in accordance with the demand for steam power through adjustment of the reactivity. An increase in power demand results in decreased steam pressure, which through automatic or inherent control causes a temporary change in reactivity, thereby increasing the rate of fuel consumption, and hence the power, until the steam pressure returns to its normal value and the power generated equals the power demanded.

A useful point of view is to consider the reactor as a source of heat that replaces the fuel input to the boiler in a central-station plant. Automatic or inherent control of its rate of heat generation thus is needed to match the fuel-feed control function in the conventional plant, a function that acts to make the power delivered equal to that demanded by the boiler and load.

Methods for Changing Reactivity

In a nuclear reactor, the excess reactivity is the independent variable through which control is effected. It determines the relative rate of increase in neutron concentration, and hence the rate of increase of power level. Thus a change in power level is related to the time integral of the excess reactivity that causes the change. The output power is the time integral of the input, excess reactivity. Consequently for a step of excess reactivity, the power builds up indefinitely with time. In this respect the reactor is a dynamically unstable integrator, or velocity device. Feedback control, either

---

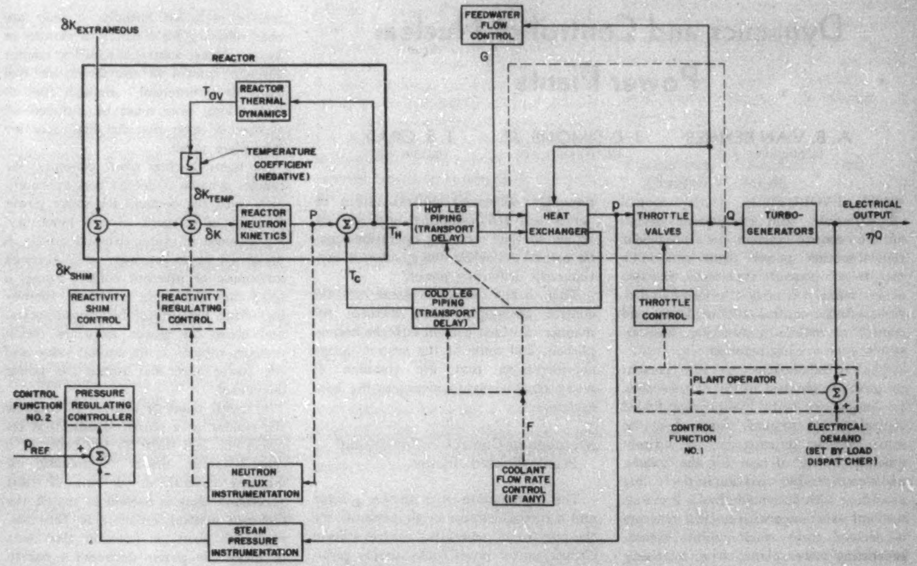
Paper 57-390, recommended by the AIEE Nucleonics Committee and approved by the AIEE Technical Operations Department for presentation at the Engineers' Joint Council Nuclear Engineering and Science Congress, Cleveland, Ohio, December 14, 1955, and the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted September 4, 1956; made available for printing January 31, 1957.

A. B. Van Rennes is with the Bendix Aviation Corporation, Detroit, Mich.; J. C. Simons, Jr. is with National Research Corporation, Cambridge, Mass.; and T. S. Gray is with the Massachusetts Institute of Technology, Cambridge, Mass.

The work described was performed as part of Project Dynamo, a study sponsored by the Atomic Energy Commission at the Massachusetts Institute of Technology.

external or internal, is needed to transform it into a stable proportional device.

Adjustment of reactivity can be made in many ways, such as by the insertion or withdrawal of fuel or absorbing material, either in a solid (rod) or liquid (concentration) form; by a change in aspect of a reflector or moderating element; or by a change in one of the physical parameters of the reactor, such as density or temperature. Classically, the position of neutron-absorbing (boron, hafnium, or cadmium) control rods has been the means for adjusting reactivity in a thermal reactor. In research and production reactors it has been usual practice to position these rods so as to maintain a constant (or otherwise specified) neutron concentration at one or more neutron flux detectors within the reactor. In a central-station power plant, in which the reactor is closely coupled to a turbogenerator, the power demand of the turbine determines the operating conditions of the reactor. This control is often exerted through the intermediate effect of the steam pressure, a dependent variable, which acts in turn to position control rods or to effect other means of adjusting the reactivity.



δK EXTRANEOUS
REACTOR
T_{av}
REACTOR THERMAL DYNAMICS
ζ
TEMPERATURE COEFFICIENT (NEGATIVE)
δK TEMP
Σ
δK
REACTOR NEUTRON KINETICS
P
T_H
T_C
δK SHIM
REACTIVITY SHIM CONTROL
REACTIVITY REGULATING CONTROL
CONTROL FUNCTION NO. 2
PRESSURE REGULATING CONTROLLER
P_{REF}
+
-
NEUTRON FLUX INSTRUMENTATION
STEAM PRESSURE INSTRUMENTATION
FEEDWATER FLOW CONTROL
G
HOT LEG PIPING (TRANSPORT DELAY)
COLD LEG PIPING (TRANSPORT DELAY)
HEAT EXCHANGER
P
THROTTLE VALVES
Q
TURBO-GENERATORS
ELECTRICAL OUTPUT
ηQ
THROTTLE CONTROL
PLANT OPERATOR
F
COOLANT FLOW RATE CONTROL (IF ANY)
CONTROL FUNCTION NO. 1
ELECTRICAL DEMAND (SET BY LOAD DISPATCHER)

Fig. 1. Block diagram of power plant including feedback control loops

Stabilization of the Reactor

The central problems to which reactor control engineers have given most of their attention are the regulation of the reactor steady power level against random changes in reactivity introduced from extraneous sources; the nature of the transient response of the power level and reactor temperature to accidental or deliberate reactivity disturbances; and control of the power level in accordance with power demand. The technique of regulation is to provide feedback to the input from the output of the device in such a way that changes in the output tend to be opposed by the effect of the resulting change in the input. The resultant system then is self-regulating.

A commonly used method for stabilization of power level is to measure neutron flux with a device such as an ionization chamber, and to employ the resulting electric signals to position a neutron-absorbing control rod. The neutron flux at the neutron detector, and hence the power level, is maintained constant. This method has been used extensively, but for effective control of reactors with short mean neutron lifetimes it requires a relatively high speed of response and considerable rod-driving power. The resulting equipment tends to be unreliable and its behavior upon failure can be dangerous to the plant.

Reactors having temperature coefficients of reactivity that are negative possess the desired feature of inherent self-regulation.[1] The negative temperature coefficient may be considered to constitute part of an internal feedback loop within the reactor which tends to maintain the average reactor temperature constant,[2] as is shown in the heavy rectangle marked reactor at the upper left of Fig. 1. Separate additional means for regulating the neutron flux and power output through neutron-flux instrumentation is illustrated as a loop, shown by a dashed line, below the rectangle.

If all other parameters affecting reactivity such as control-rod position, fuel, and poison content are fixed, the reactor having a negative temperature coefficient of reactivity is critical only at a particular temperature. At any other temperature the power developed is not constant but tends to rise or fall. In response to a

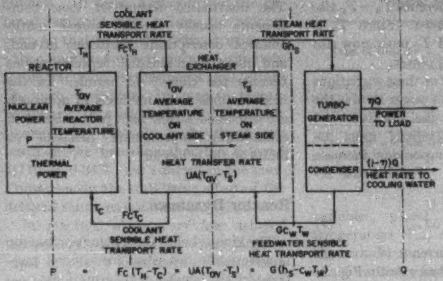

COOLANT SENSIBLE HEAT TRANSPORT RATE
FcT_H
T_H
REACTOR
NUCLEAR POWER
T_{av} AVERAGE REACTOR TEMPERATURE
P
THERMAL POWER
T_C
FcT_C
COOLANT SENSIBLE HEAT TRANSPORT RATE
HEAT EXCHANGER
T_{av} AVERAGE TEMPERATURE ON COOLANT SIDE
T_S AVERAGE TEMPERATURE ON STEAM SIDE
HEAT TRANSFER RATE $UA(T_{av}-T_S)$
STEAM HEAT TRANSPORT RATE
Gh_S
Gc_wT_w
FEEDWATER SENSIBLE HEAT TRANSPORT RATE
TURBO-GENERATOR
CONDENSER
ηQ POWER TO LOAD
$(1-\eta)Q$ HEAT RATE TO COOLING WATER
$P = Fc(T_H-T_C) = UA(T_{av}-T_S) = G(h_S-c_wT_w)$
Q

Fig. 2. Schematic diagram of energy flow in steady state

change in power demand, the surplus or deficiency of power generated causes the reactor temperature temporarily to rise or fall. This temperature change causes the reactivity first to fall or rise correspondingly. The power generated, in turn, falls or rises beyond the newly demanded value, causing the temperature to return to the value corresponding to criticality at the same time that the power generated reaches equality with the power demanded. Any such power transient must end with eventual return to the same temperature, but at the new required power level. A variation in reactivity due to some other cause will result in establishment of a different temperature at which criticality is again attained.

Nomenclature

P = reactor power
F = coolant mass flow rate
c = specific heat of coolant
T_H = temperature of coolant leaving reactor
T_C = temperature of coolant entering reactor
U = over-all heat transfer coefficient of heat exchanger
A = heat transfer area of heat exchanger
T_{av} = average coolant temperature in reactor = average coolant temperature in heat exchanger
T_s = steam temperature
G = mass flow rate of steam = mass flow rate of feedwater
h_s = heat content of steam per unit mass
c_w = specific heat of feedwater
T_w = temperature of feedwater
Q = power absorbed by turbines and condensers
η = thermal and mechanical efficiency of turbogenerator
ηQ = electric power delivered to bus bars
C_i = concentration of ith delayed emitter (six in all)
n = neutron concentration
δK = increment in reactivity
λ_i = decay constant of ith delayed emitter
$\beta = \Sigma\beta_i$ = fractional contribution to total neutron population by all delayed emitters
β_i = fractional contribution to total neutron population by ith delayed emitter
l = neutron lifetime
S = contribution to rate of increase of neutron concentration from neutron source within reactor
M' = mass of total coolant in reactor, heat exchanger, and coolant loop

Static Plant Control

The primary control of any central-station plant lies with the load dispatcher, who orders a change in the electrical output delivered to the bus bars by the plant in response to the demands of the power distribution network. The plant operator adjusts his electrical output according to the load dispatcher's orders by varying the steam flow to the turbines, as indicated in Fig. 1.

The identification of electrical demand as the primary control signal for a nuclear central-station plant is the major distinction from a research or production reactor, since control of these latter types has always been based on maintaining the neutron flux density at a level prescribed by requirements of an experiment or by production considerations.

With electrical demand established as the primary control, the variations of other plant variables with power output must be examined to determine whether the resulting ranges of values are compatible with the limitations of equipment and materials.

For those reactors stabilized by a negative temperature coefficient ζ, the average temperature within the reactor T_{av} tends to be maintained at a constant value, independent of the rate of energy released within the reactor P, unless external variations of reactivity occur. The dependence of other variables in the plant on P can be found by consideration of the pertinent energy balances. In the static condition (all variables constant in time), these rates of energy transfer must be equal:

1. The power generation in the reactor.
2. The net rate of heat transport by the primary coolant.
3. The heat transfer rate across the heat exchanger tubes.
4. The net rate of heat transport by the steam and feedwater.
5. The rate of energy absorption by turbines and condensers.

Fig. 2 is a schematic description of the several rates of energy flow. The equivalent mathematical statement is

$$P = Fc(T_H - T_C) = UA(T_{av} - T_s) = G(h_s - c_w T_w) = Q \quad (1)$$

The analytical expressions evidently neglect intermediate heat losses, and the temperatures indicated may not be measurable. Nevertheless the picture is one which adequately describes the process and from which some characteristics of the plant may easily be deduced. It is convenient and reasonable to make the additional assumption

$$T_{av} = \frac{1}{2}(T_H + T_C) \quad (2)$$

For simplicity and reliability, and to avoid the complications in dynamic behavior that would otherwise result, F is usually held constant.

If in those reactors having a significant temperature coefficient, T_{av} is maintained constant, equations 1 and 2 force T_H and T_C to follow the pattern of Fig. 3, since $T_H - T_C$ must be proportional to power P. Since U does not change greatly with operating conditions, $T_{av} - T_s$ also is proportional to power. T_s is very nearly the saturation temperature of the steam leaving the boiler, and so the steam pressure is determined as the saturation value p corresponding to T_s. The temperature of the steam delivered to the turbine may be increased above T_s by superheaters, but their use does not materially affect the energy flow pattern.

Fig. 3 indicates that in a temperature-regulated reactor, as the pressure p, temperature T_s, and the value of h_s corresponding to T_s drop with increasing power P, the steam flow rate G must increase with P at a rate faster than proportional. Thus with T_{av} constant; i.e., without any external change of reactivity, the

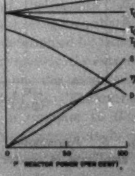


Fig. 3. Static plant characteristics with T_{av} constant

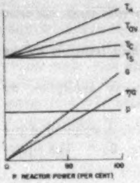


Fig. 4. Static plant characteristics with constant steam pressure

desired electric load ηQ may be supplied simply by proper adjustment of G; in other words, by adjustment of the steam throttle. The steam pressure p varies with G.

Hence, assuming a dynamically stable plant, variations in the electric load demand can be met simply by changing the steam throttle, without changing the external reactivity adjustment. However, steam pressure falls rapidly with increasing load, and so does plant efficiency. If steam pressure and plant efficiency are established by the full-load requirements, the steam pressure rises considerably higher at low loads. Steam plant equipment and auxiliaries, commonly designed to operate at constant steam pressure, do not perform well at other steam pressures, and actually may fail to operate, e.g., if the steam pressure exceeds the maximum head of the feedwater pumps.

To achieve optimum plant performance, it is therefore necessary to impose an additional restriction on operation, namely that the steam pressure be maintained essentially constant, independent of power output.

The consequence of this restriction can be determined from equation 1. Specifying the pressure p also fixes T_s and h_s; thus the net heat transport rate to the turbine and condenser per unit mass of steam, $h_s - c_w T_w$, is constant. Steam flow G and $T_{av} - T_s$ thus vary linearly with power as shown in Fig. 4.

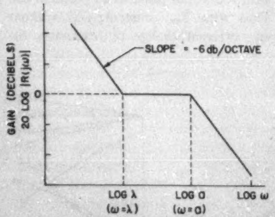


Fig. 5. Asymptotic approximation for amplitude of reactor transfer function

The temperature difference $T_H - T_C$ also is proportional to power. Since T_s is constant, T_{av}, T_H, and T_C must now rise linearly with increasing power.

The most important of these variations caused by an increase of power level is the increase of T_{av}, since the reactivity reduction resulting therefrom must be cancelled by an accompanying increase of external reactivity. The increase required (see Fig. 1) is

$$\delta K_{\text{shim}} = -\delta K_{\text{temp}} = -\zeta(\Delta T_{\text{av}}) \qquad (3)$$

where ΔT_{av} is the increase of average temperature in accordance with Fig. 4, and the numerical value for ζ is negative. This adjustment in reactivity is the adjustment by which steam pressure is maintained constant, and is most readily made by a feedback control loop which senses changes in steam pressure. Thus, in addition to the primary control function, that of matching electrical demand, a secondary control loop maintains steam pressure at the desired value by adjustment of reactivity, as shown in Fig. 1. This secondary control function may be either manual or automatic.

Block Diagram of Entire System

The block diagram of Fig. 1 represents one concept of a complete power plant, including the two control loops that govern the plant in the operating or power range (10% to 100% full power). The blocks are by no means restricted to linear components. The diagram is intended to represent the complex plant by the simplest description that adequately shows the flow of energy and information between components important to the static and dynamic behavior of the plant. The supervisory or safety circuits are not shown, since they act only if abnormalities develop which cause any of the monitored variables to exceed preset limits. With minor exceptions or revisions the diagram can correctly represent many nuclear powered central-station systems.

Dynamic Behavior of Total Plant

Whether stabilization is accomplished by means of a negative temperature coefficient or by an external control system, a feedback system exists whose dynamic stability must be examined under the various possible operating conditions. Analytical determination of the dynamic stability of a reactor and its associated power plant is extremely difficult since the dynamic thermal characteristics of the reactor, heat exchangers, and coolant piping interact strongly with each other. The distributed nature of some components usually renders simple mathematical descriptions completely invalid, and more sophisticated representations become too clumsy to handle. In the next sections some methods of approach are discussed which may prove useful in establishing approximate behavior or limits on the behavior.

Reactor Dynamics

The kinetic behavior of a nuclear reactor is adequately described for control purposes by the familiar set of equations:

$$\frac{dn}{dt} = \frac{\delta K - \beta}{l} n + \sum_{i=1}^{6} \lambda_i C_i + S \qquad (4A)$$

$$\frac{dC_i}{dt} = \frac{\beta_i}{l} n - \lambda_i C_i \qquad (4B)$$

For practical application, this seventh-order description is unwieldy, and so several approximate methods for expressing the reactor characteristics are used, depending on the purpose of the analysis. For studying small-amplitude excursions about an operating point, a linearized representation is appropriate. Such representations for conventional *U-235* reactors (in which 0.76% of the neutrons are delayed) have been described elsewhere.[3] Within moderate relative power ranges a simplified Laplace transfer function of the form[4]

$$R(s) = \frac{n(s)}{\delta K(s)} = \frac{n(s+\lambda)}{ls(s+a)} \text{ in which } a > \lambda \qquad (5)$$

is usually adequate.

An asymptote approximation[5] to the amplitude of the transfer function $R(j\omega)$ from equation 5, (on decibel versus log ω scales) is shown in Fig. 5. It reveals that the reactor behaves approximately as an integrator but with decreased gain at low frequency, because of the effect of delayed neutrons.

In *U-233* reactors, the form of the representation is the same, but the fraction of neutrons delayed is only 0.24% and the relative concentration of the characteristic delayed emitters is different, so that the separation between λ and a on the frequency scale is reduced. The stabilizing effect of the delayed neutrons is less and a smaller positive reactivity excess can make the reactor prompt critical. Likewise, in reactors ultimately fueled partially with *Pu-239* (either through loading or through conversion), the fewer delayed neutrons result similarly in a modified system function.

A further complication arises when reactors involving circulating fuel are studied. In these, a majority of the delayed neutrons exist outside the reactor core, so that of all the delayed neutrons only a fraction proportional approximately to the fractional fuel residence time within the core contribute to reactor kinetic behavior. In a particular *U-233* reactor concept, for example, only about one tenth of the neutrons that are delayed (1/10 of 0.24%) are emitted in the core. Consequently the stabilizing effect of the delayed neutrons is almost negligible.

In reactors where very few delayed neutrons are available, or where large or rapid reactivity excursions are being contemplated, the contribution of the delayed neutrons is often neglected; i.e., prompt criticality is assumed. The simplified kinetic equation for a reactor under these conditions

$$\frac{dn}{dt}=\frac{\delta K}{l}n \qquad (6)$$

can also be written

$$\frac{dP}{dt}=\frac{\delta K}{l}P \qquad (7)$$

since the power released is proportional to the rate of fission, which in turn is proportional to the neutron concentration. If the quantity P on the right-hand side of equation 7 is assumed constant at a value P_0, the Laplace transfer function for the incremental behavior of the reactor is found by the usual methods to be

$$R(s)=\frac{P(s)}{\delta K(s)}=\frac{P_0}{ls} \qquad (8)$$

This transfer function is useful in linear incremental analyses only for relatively small fluctuations in P. The reactor characteristic is then like that of an integrator, but with its coefficient of proportionality proportional to the quiescent power level. For larger ranges of variation, the nonlinear differential equations, equations 4, of the system must be considered directly,[6] or a system analysis based upon logarithmic functions of flux and power must be used. A set of relevant kinetic parameters for four different reactor concepts is presented in Table I.

Table I. Kinetic Parameters of Representative Reactor Concepts

Type of Reactor	Fuel	Neutron Lifetime (Seconds)	$\beta=\Sigma\beta_i$	$\beta_{eff}=\Sigma\epsilon_i\beta_i$*	$\zeta(\delta K/C)$†
Pressurized H_2O	*U-235*	0.00003	0.0076	0.0076	−0.00030
Na-Graphite	*U-235*	0.0006	0.0076	0.0076	−0.00002
Liquid-metal fuel	*U-233*	0.007	0.0024	0.00027	−0.00030
Aqueous homogeneous	*U-233*	0.0017	0.0024	0.00057	−0.0011

* The term ϵ_i = that fraction of each delayed neutron component which contributes to the delayed neutron population in the core.

† Degrees centigrade.

Short-Term Dynamics With Temperature Feedback

The foregoing discussion considers the excess reactivity as an independent variable. Actually the reactor is part of a dynamic feedback system composed of the reactor neutron kinetics, reactor thermal characteristics, and the negative temperature coefficient of reactivity as shown within the reactor block shown in Fig. 1. The feedback loop relates the significant variables, power P_0, average reactor temperature T_{av}, and excess reactivity δK, as indicated; and δK depends on the other variables because of the feedback. The system can be represented in a quasi-stationary manner by the equations

$$P=Fc(T_H-T_C)+Mc\frac{dT_{av}}{dt} \qquad (9)$$

$$\delta K=\zeta(T_{av}-T_{ref}) \qquad (10)$$

together with equations 4(A) and 4(B) which describe the neutron kinetics. Here M is the mass of the coolant within the reactor and the other symbols have the same meanings as before. Actually the temperature contribution to the reactivity is determined by the integrated weighted temperature distribution throughout the reactor, but for purposes of analysis it is convenient to assume that a mean temperature T_{av} can be defined, in terms of which the reactivity dependence on temperature is given by equation 10. For this quasi-stationary analysis, T_C is considered an independent variable. Any dependence of T_C on T_H (via the heat exchanger) is, for the moment, neglected.

If the further assumption of equation 2

$$T_{av}=\frac{1}{2}(T_H+T_C)$$

is made, equation 9 becomes

$$P=2Fc(T_{av}-T_C)+Mc\frac{dT_{av}}{dt} \qquad (11)$$

which yields the transformed relationship

$$T_{av}(s)=\frac{1}{(1+\tau s)}\left[\frac{P(s)}{2Fc}+T_C(s)\right] \qquad (12)$$

Thus the transfer function of the thermal load is

$$H(s)=\frac{T_{av}(s)}{P(s)}=\frac{\frac{1}{2Fc}}{(1+\tau s)} \qquad (13)$$

in which the time constant τ is defined as $M/2F$ and is one half the transit time of the coolant through the reactor. Equation 13 indicates that in the time domain the response of the average reactor temperature T_{av}, to a unit step in nuclear power P, is a rising exponential having an amplitude $1/(2Fc)$ and a time constant τ. The block-diagram equivalent of the system is given in Fig. 6 with the reactor kinetics represented simply by equation 8. Recall that the symbols in the above equations represent increments in the variables.

The degree of stability of the system of Fig. 6 can now be examined. The stability of a feedback system describable by linear equations is evaluated by plotting on the complex plane as a function of frequency the locus of the loop transmission or loop gain; i.e., of the product of the transfer functions around the loop. Thus, if the reactor is characterized by equation 8 and the heat transfer mechanism by equation 13, the loop transmission is the product locus $\zeta R(j\omega)H(j\omega)$. This product locus appears as shown in solid lines in Fig. 7.

The loci in Fig. 7 shown by dashed lines are those resulting when the effect of delayed neutrons is significant, so that the reactor must be characterized by equation 5. The proximity of the $\zeta R(j\omega)H(j\omega)$ locus to the point $(-1,0)$ determines the degree of stability; for the system to oscillate in an undamped manner, the locus must pass through or encircle this critical point. Because the transfer function $R(s)$ of a reactor alone is proportional to the power level, the loop gain of this feedback system is likewise proportional to the power level. According to this model, however, the system is never capable of steady oscillation, because the effect of an increase in loop gain is only to bring the product locus into closer proximity to the unstable point, regardless of the parameter values in the transfer functions $H(j\omega)$ and $R(j\omega)$. On the other hand, it must be recognized that a transient disturbance can cause damped oscillations or severe overshoots if the loop gain is reasonably large.

From a study of the reactor block diagram in Fig. 1, one recognizes that the reactor behavior is determined by two independent variables; namely, the extraneous reactivity, δK_{ext}, and the temperature of the entering coolant T_C. The over-all transfer functions in response to these two excitations relate increments in reactor power $P(s)$, and average temperature

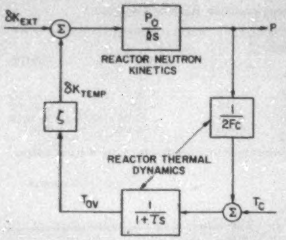


Fig. 6. Feedback loop that provides reactor stabilization by negative temperature coefficient

$T_{av}(s)$, to increments in $\delta K_{ext}(s)$ and $T_C(s)$, in the following manner

$$\frac{P(s)}{\delta K_{ext}(s)}=\frac{P_0(1+\tau s)}{ls(1+\tau s)-\dfrac{P_0\zeta}{(2Fc)}}$$

$$=\frac{\dfrac{2P_0}{-\zeta\Delta T}(1+\tau s)}{\left(1+\dfrac{2\gamma}{\omega_n}s+\dfrac{s^2}{\omega_n^2}\right)} \qquad (14)$$

$$\frac{T_{av}(s)}{\delta K_{ext}(s)}=\frac{\dfrac{P_0}{2Fc}}{ls(1+\tau s)-\dfrac{P_0\zeta}{(2Fc)}}$$

$$=\frac{\dfrac{1}{-\zeta}}{\left(1+\dfrac{2\gamma}{\omega_n}s+\dfrac{s^2}{\omega_n^2}\right)} \qquad (15)$$

$$\frac{P(s)}{T_C(s)}=\frac{P_0\zeta}{ls(1+\tau s)-\dfrac{P_0\zeta}{(2Fc)}}=\frac{-\left(\dfrac{2P_0}{\Delta T}\right)}{\left(1+\dfrac{2\gamma}{\omega_n}s+\dfrac{s^2}{\omega_n^2}\right)} \qquad (16)$$

$$\frac{T_{av}(s)}{T_C(s)}=\frac{ls}{ls(1+\tau s)-\dfrac{P_0\zeta}{(2Fc)}}=\frac{\left(\dfrac{2l}{-\zeta\Delta T}\right)s}{\left(1+\dfrac{2\gamma}{\omega_n}s+\dfrac{s^2}{\omega_n^2}\right)} \qquad (17)$$

where the substitutions

$$\omega_n^2=\frac{-\zeta P_0}{2lFc\tau}=\frac{-\zeta\Delta T}{2l\tau} \qquad (18A)$$

$$\gamma\omega_n=\frac{1}{2\tau} \qquad (18B)$$

$$\gamma=\sqrt{\frac{lFc}{-2\zeta P_0\tau}}=\sqrt{\frac{l}{-2\zeta\Delta T\tau}} \qquad (18C)$$

$$\Delta T=\frac{P_0}{Fc}=(T_H-T_C) \qquad (18D)$$

have been introduced. (Note that the definition of ζ includes the sign of the temperature coefficient, so that for negative temperature coefficients $-\zeta$ is a positive quantity and ω_n is a real number.)

These transfer functions are useful in interpreting the various features of the transient response of the reactor to disturbances of δK_{ext} or T_C. Equation 14, for example, when compared with the simple transfer function of a reactor only, equation 8, verifies that the effect of temperature feedback is to convert the integrating characteristic of a reactor to one essentially proportional in nature. The characteristics of chief interest are maximum amplitude, rate of attenuation, oscillation frequency, and final steady-state value. Final values in response to unit step excitations can be obtained from the static plant relations, but they can be found more readily from the transfer functions by setting s equal to zero in equations 14 to 17. It is not difficult in any of these transfer functions to show that the peak amplitude of the response to a step input never exceeds twice the final deviation of the dependent variable from its initial value. (This fact has previously been stated as Nordheim's accident theorem.) Since $P_0/\Delta T$ $(=Fc)$ is a constant independent of P_0, the peak amplitude of the response to a δK_{ext} step depends only on the value of the step and not on the operating power level. The frequency and damping will change, however, with power level.

The model assumed in the foregoing does not include any transport time lags. Additional phase shifts may well be present so that instability is still a possible consequence of excessive internal gain. To ensure stability in view of additional phase shift, it is conservative to require at least a 40-degree phase margin at full power, which corresponds for the quadratic behavior to a damping ratio γ of 0.4. If the power should increase to four times full power, γ would drop only to 0.2 and the phase margin would be reduced to 25 degrees, which is still likely to be adequate. At 10% full power γ becomes approximately 1.2, which is quite satisfactory. This stability criterion can be expressed as

$$\sqrt{\frac{lFc}{-2\zeta P_0\tau}}\geq 0.4 \qquad (19)$$

where P_0 is the full-power value. Alternatively one can write

$$0.32(-\zeta)P_0\tau\leq Fcl \qquad (20)$$

which shows that there is a maximum limit on the product $(-\zeta)P_0\tau$. For a specified coolant flow rate F and transport time 2τ in a particular reactor, equation 20 imposes an upper limit on the value of power level P_0. Exceeding this limit can cause not only highly underdamped behavior but, if additional phase lags exist, can actually lead to a growing oscillatory behavior. Linear incremental analysis is inadequate fully to explore the possibility of runaway behavior in a reactor-powered plant; the application of nonlinear dynamics to the concept of stability in nuclear systems has been studied.[6–8]

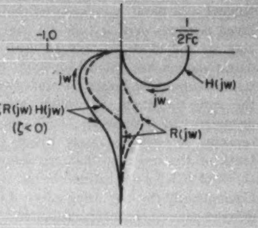


Fig. 7. Locus diagram of internal loop transmission within reactor

If γ is never less than 0.4 the transient response will never exhibit more than two significant overshoots. It is also interesting to examine the quantity $\gamma\omega_n$ which determines the rate of attenuation of any transient. From equation 18(B) one sees that $\gamma\omega_n$ has the value $1/(2\tau)$ regardless of power level. This means, for example, that after a time interval 4τ (twice the transit time through the reactor) a transient amplitude will be attenuated to within 15% of its final incremental value. This result has desirable significance, for it indicates that transients within the reactor are damped out in times quite short compared with the coolant loop circulation time; thus any resonance interaction between the reactor and the coolant loop tends to be filtered by the heat exchanger.

This treatment of the dynamic behavior within the reactor vessel has several limitations. It only sets forth the relationship between increments in the significant variables. It is not a detailed description of the internal mechanism of the reactor, although it presents a reasonable external picture. Furthermore, it holds true only for time intervals less than the coolant circulation time, 2τ, since it was tacitly assumed that T_C was independent of the previous history of T_H. In practice, the circulating coolant provides dynamic feedback from T_H to T_C, thereby adding additional complication to the model over longer time intervals.

Long-Term Dynamics With Temperature Feedback

For time intervals long compared to the coolant circulation time, in other words, for very slow transients, the behavior of the reactor, coolant loop, and heat exchanger depends on a different set of parameters. Assume that all of the circulating coolant has an average temperature T_{av}, which is the temperature governing the rate of heat transfer to water and steam at a temperature T_s. The following relation then holds:

$$P=UA(T_{av}-T_s)+M'c\frac{dT_{av}}{dt} \tag{21}$$

This expression considers the entire coolant system to be essentially a lumped element; i.e., it neglects the effect of transport delays of the circulating coolant. The equivalent transform relation is

$$T_{av}(s)=\frac{1}{(\tau's+1)}\left[\frac{P(s)}{UA}+T_s(s)\right] \tag{22}$$

where

$$\tau'=\frac{M'c}{UA}$$

This relation of T_{av} to P is seen to be similar to equation 12, not only in form but also in that it is an alternate description of the thermal dynamics of the reactor. Consequently, equation 22 together with equations 8 and 10 describe a stabilizing feedback loop around the reactor. The equivalent block diagram is the same as that of Fig. 6 with $1/(2Fc)$ replaced by $1/(UA)$, τ by τ', and T_C by T_s. The closed-loop transfer functions relating the dependent variables T_{av} and P to the independent variable T_s are

$$\frac{T_{av}(s)}{T_s(s)}=\frac{l's}{l's(1+\tau's)-\frac{P_0\zeta}{UA}}=\frac{\left(\frac{l'}{-\zeta\Delta T'}\right)s}{\left(1+\frac{2\gamma}{\omega_n}s+\frac{s^2}{\omega_n^2}\right)} \tag{23}$$

$$\frac{P(s)}{T_s(s)}=\frac{P_0\zeta}{l's(1+\tau's)-\frac{P_0\zeta}{AU}}=\frac{-\left(\frac{P_0}{\Delta T'}\right)}{\left(1+\frac{2\gamma}{\omega_n}s+\frac{s^2}{\omega_n^2}\right)} \tag{24}$$

where

$$l'=\frac{a}{\lambda}l \text{ (see equation 5, at low frequencies)} \tag{25A}$$

$$\Delta T'=\frac{P_0}{UA}=(T_{av}-T_s) \text{ corresponding to } P_0 \tag{25B}$$

$$\omega_n^2=\frac{-\zeta P_0}{UAl'\tau'}=\frac{-\zeta\Delta T'}{l'\tau'} \tag{25C}$$

$$\gamma\omega_n=\frac{1}{2\tau'} \tag{25D}$$

$$\gamma=\frac{1}{2}\sqrt{\frac{l'UA}{-\zeta P_0\tau'}}=\frac{1}{2}\sqrt{\frac{l'}{-\zeta\Delta T'\tau'}} \tag{25E}$$

The time constant τ' is related to the coolant circulation time t_c by

$$\tau'=\frac{\Delta T'}{\Delta T}t_c \tag{26}$$

Thus τ' and t_c are of the same order of magnitude. The attenuation factor $\gamma\omega_n$ is equal to $1/(2\tau')$; hence any transient is reduced to within 15% of its final increment value in a time $4\tau'$, which is quite acceptable. Since excessively underdamped behavior is undesirable, the damping ratio γ should be at least 0.3, which establishes the additional criterion

$$0.36(-\zeta)P_0\tau'\leq l'UA \tag{27}$$

Equation 27, in a manner similar to equation 20, establishes a limitation on reactor power P if excessively underdamped behavior or growing oscillations are to be avoided. Whereas from equation 20

$$P_0\leq\frac{3.1lFc}{-\zeta\tau} \tag{28}$$

equation 27 imposes the constraint

$$P_0\leq\frac{2.8\left(\frac{a}{\lambda}\right)lUA}{-\zeta\tau'} \tag{29}$$

Whether the short-term or long-term constraint is more restrictive depends on the relative values of the various parameters. (In general, $\tau'>>\tau$ and $a>\lambda$.) This treatment, valid only for time intervals which are long compared with the coolant circulation time, presents an external picture of the incremental behavior of plant variables. It does not include the fact that T_s and T_{av} are interrelated by the thermal dynamics of the secondary coolant loop and steam turbine, dynamics which are difficult indeed to describe.

Summary

A study of the steady-state energy pattern in a nuclear power plant as a function of power level is necessary for proper synthesis of a complete system. Nuclear-fueled plants differ from conventional plants chiefly in that the heat source alone has the characteristics of an integrator, i.e., a characteristic in which the time rate of change of power level depends on the control variable, excess reactivity. The existence of a negative temperature coefficient of reactivity, if present, constitutes an internal negative feedback loop which transforms the reactor characteristic into that of a proportional device. In the absence of other disturbance or control, this feedback dictates that the average reactor temperature be independent of power level. A deliberate external control loop is necessary if steam pressure, steam temperatue, or another appropriate plant variable is chosen to be held constant.

Furthermore, study of the dynamic behavior of a nuclear power plant in which the reactor is closely coupled to a turbine through a heat exchanger is essential to ensure satisfactorily stable performance. Dynamic analyses of the entire system are made first on the assumption that the transport time of the coolant through the reactor is long compared with the time interval of interest, and second on the opposite assumption. These analyses lead to expressions given that aid in evaluation of such characteristics as the maximum amplitude, rate of attenuation, oscillation frequency, and final steady-state values of significant variables in the system.

References

1. Nuclear Power Plant Control Considerations, M. A. Schultz. *AIEE Transactions*, vol. 72, pt. I, Mar. 1953, pp. 22–26.

2. Synthesis of Control for Nuclear Power Plants, J. N. Grace. *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 9, 1954, pp. 83–85.

3. Pile Transfer Functions, J. P. Franz. Atomic Energy Commission document, *AECD-3260*, Office of Technical Services, U. S. Department of Commerce, Washington, D. C., July 18, 1949.

4. Automatic Control of Power Reactors, M. A. Schultz. Atomic Energy Commission document, *AECD-3163*, *Ibid.*, Nov. 6, 1950.

5. Transfer Function of Argonne CP-2 Reactor, J. M. Harrer, R. E. Boyar, D. Krucoff. *Nucleonics*, New York, N. Y., vol. 10, no. 8, 1952, pp. 32–36.

6. The Dependence of Reactor Kinetics on Temperature, J. Chernick. *BNL 173*, Brookhaven National Laboratory, Upton, N. Y., Dec. 20, 1951.

7. Concept of Stability for Nuclear Reactors, L. B. Robinson. *Journal of Applied Physics*, New York, N. Y., vol. 25, no. 4, 1954, pp. 516–18.

8. Some Aspects of Nonlinear Dynamics, W. K. Ergen, A. M. Weinberg. *Physica*, Amsterdam, Netherlands, vol. 20, 1954, pp. 413–26.

A New Family of Transistor Switching Circuits

MORRIS RUBINOFF
ASSOCIATE MEMBER AIEE

OVER THE PAST 10 years, electronic switching circuits have found their way into numerous pieces of equipment in many fields of endeavor. The circuits are carefully designed prototypes intended to serve with utmost reliability and in large numbers in serial or parallel digital computers, in control systems, in communications equipment or in dial telephone switching offices.[1-4] The specific prototype circuit may be designed around vacuum tubes, diodes, magnetic devices, or transistors, as influenced by the system in which it is to be used. The existence of numerous basically different prototype circuits is indicative of the large variation in system goals; new families of switching circuits offer certain advantages in existing systems and open the door to new system applications. The family to be described in this paper is a new prototype available to the system designer.

In a previous paper,[5] the author and his colleagues at Philco Corporation introduced a family of direct-coupled transistor switching circuits, now commonly referred to as DCTL. These circuits have formed the basis of a number of parallel digital computers and appear to give reliable performance with great economy in total computer component count. Experience with DCTL circuits has, as usual, brought to light certain limitations which take on importance in other applications where, for example, the ultimate in switching speed is a basic requirement, or where, as in telephone offices, the quiescent or stand-by power dissipation of a large system is of considerable importance.

This paper introduces a new family of "dual-range" transistor switching circuits (circuits characterized by two pairs of output levels), which uses both n-p-n and p-n-p transistors. The dual-range family uses the same number of transistors as DCTL but requires many more capacitors and resistors, three power supplies instead of one, and several times as much power dissipation. On the other hand, the dual-range family offers a number of advantages over DCTL and other transistor switching circuits which may be summarized as follows:

1. OFF transistors are fully cut off, except for the irreducible leakage I_{c0}.
2. The base current of every ON transistor is controlled to a narrow tolerance.
3. The full common-emitter current gain is available to the designer.
4. Fast switching can be achieved by capacitor shunts across the coupling resistors.
5. Sneak paths cannot occur from base through collector.
6. A system can be designed so that almost no current is flowing in the stand-by state.

The reasons for these advantages should become clear from the following discussion.

Description of Dual-Range Family

The simplest dual-range circuit is the flip-flop, shown in Fig. 1. The basic flip-flop is an asymmetrical both-on or both-off circuit using one p-n-p and one n-p-n transistor. In the OFF state, each transistor is held off by bias applied to its base through load and transfer resistors designated R_L and R_T. In the ON state, each transistor is held on by base current provided through transfer resistor R_T in series with the collector-emitter circuit of the opposite transistor.

Four supply voltages are provided; in order of decreasing voltage they are V_{P+}, V_{N+}, V_{P-}, and V_{N-}. Only three physical supplies are needed since any one of these can of course be chosen as system ground. Two pairs of output levels are provided by the flip-flop, one at each collector. The n-p-n collector is approximately at voltage V_{P+} when the flip-flop is off; this will be referred to as the $P+$ level. In the ON state the n-p-n collector dwells at the $P-$ level, with voltage slightly over V_{P-}. Similarly the p-n-p collector is at the $N-$ level when OFF and at the $N+$ level when ON.

Since both transistors are cut off in the OFF state, it might be suspected that a system incorporating these flip-flops could be designed so that all flip-flops are nonconductive in the stand-by state (item 6 in the foregoing list of advantages). That such a system with minimum stand-by current can in fact be designed, as regards both flip-flops and switching circuits, is the subject of another paper currently in preparation by H. E. Tompkins. This advantage of the dual-range family is important in systems having a large stand-by to operating ratio, such as telephone offices: it is not as important in continuously operating systems such as digital computers. A disadvantage of minimum stand-by current is the large current surges which occur during operation; this should be contrasted with the almost constant supply current of a DCTL system.

Again, when the flip-flop of Fig. 1 is conducting, almost the entire voltage drop (V_{N+} to V_{P-}) appears across each resistor R_T. The base current of each transistor is therefore determined almost entirely by its transfer resistor and the power supply voltages. This is true for all dual-range circuits and leads to closely controlled base drive (item 2 in the foregoing).

Because of the two ranges of output voltages, each line in a dual-range schematic is labelled with both its logical function name and the signal level corresponding to that function. For example, flip-flop J of Fig. 1 has its n-p-n collector labelled J, $P-$; this indicates that when the flip-flop is set ($J=1$), this collector is at the $P-$ level, i.e., at voltage slightly more positive than V_{P-}.

Note that there is electrical inversion in the two halves of the flip-flop but that this is not logical inversion. It is characteristic of all dual-range circuits that the outputs of n-p-n switching networks go negative and that those of p-n-p networks go positive when the networks go ON or are conducting. Since n-p-n networks alternate with p-n-p networks, this electrical inversion results in both being switched on (or off) simultaneously, corresponding to direct logical connection, not logical inversion.

Paper 57-105, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 22, 1956; made available for printing December 5, 1956.

MORRIS RUBINOFF is with the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa.

The author wishes to express his appreciation of suggestions and assistance provided by members of the Moore School dial telephone switching circuits project sponsored by the Signal Corps Engineering Laboratories. The term "dual-range" was suggested by H. E. Tompkins, who also developed the logical nomenclature used in this paper and the concept of using "set" and "hold" circuits to control flip-flops.

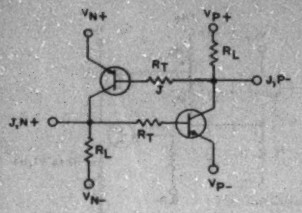


Fig. 1 (left). Basic dual-range flip-flop with ON-state J

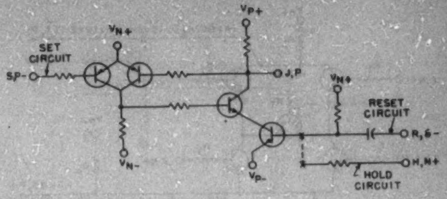


Fig. 2 (right). Set, reset, and hold circuits for dual-range flip-flop

The flip-flop may be set as shown in Fig. 2, by a set signal *S*, *P*− into the p-n-p transistor paralleling the p-n-p flip-flop transistor. A signal *S*, *N*+ could just as readily set the flip-flop through an n-p-n transistor paralleling the n-p-n flip-flop transistor. This duality in gating possibilities is characteristic of the entire dual-range family.

The flip-flop may be reset as shown in Fig. 2, by a reset signal in the form of a negative pulse *R*, δ− into the n-p-n transistor in series with the n-p-n flip-flop transistor. Or the flip-flop may be held off by the signal *H*, *N*+, which retires only when a set signal is anticipated; this mode of operation offers advantages in telephone office service.[4] Resetting and holding can also be accomplished, of course, in the p-n-p half of the flip-flop.

Both *P*-type and *N*-type dual-range combinational switching circuits are available to the designer. Fig. 3(A) shows a 2-high series-parallel *P*-type gate. Each of the four transistors is cut off in the absence of its input signal, i.e., when the signal applied to its base is at the *N*− level. Note that since V_{N-} is more negative than V_{P-}, the transistor is completely cut off except for the small collector leakage I_{c0}. This is advantage 1 of the foregoing list.

Now suppose that *A* is present and hence at *N*+ level, but *B* is absent or at *N*− level. Then the base of the *A* transistor rises only to the *N*+ level, which is still more negative than V_{P+}; this is advantage 5. In the DCTL circuit, for example, the base voltage can sometimes exceed collector voltage in a circuit of this type, with the result that unwanted base-to-collector current flows.

The circuit of Fig. 3(A) is a mixed gate-buffer or AND-OR switch. Full collector current will flow when either *A* and *B*, or *C* and *D*, or all four input signals are present. This is indicated by the logical notation (*ABvCD*, *P*−) applied to the output line. Fig. 3(B) shows an *N*-type gate of mixed height for which collector current will flow when either *A*, or *B* and *C* and *D*, or all four inputs are energized. Here the output is labelled (*AvBCD*, *N*+) indicating that the output voltage levels are *N*− and *N*+ rather than the *P*+ and *P*− of Fig. 3(A). Note also that *P*-type signals serve as inputs to *N*-type switching circuits.

More complicated combinational circuits can be obtained by alternately cascading *P*-type and *N*-type circuits. For example, the output of Fig. 3(A) might be connected to one of the inputs of Fig. 3(B), say input *A*, *P*−. In the conducting state, current flows from V_{N+} through the emitter-base diode and the transfer resistor of *A*, *P*− and then through two series collector-emitter paths to V_{P-}. Here it may be noted again, as already mentioned, that there is no logical inversion between successive networks even though electrical inversion occurs.

Note also that the base current for *A*, *P*− is determined by the base resistor, the difference in supply voltages, and the loss of voltage across the base-emitter diode and the series collector-emitters. If V_{N+} be chosen about 8 volts more positive than V_{P-}, a collector-emitter drop of 0.15 or 0.2 volt can be tolerated. In other words, saturation I_c may be made available and full advantage taken of grounded-emitter current gain, advantage 3. In DCTL circuits, for example, the collector-emitter drop must be held below 0.15 volt, which limits the available I_c.

The choice of V_{P+} (and similarly V_{N-}) is determined by the total drop due to leakage current through R_L in the OFF state and the prescribed upper limit on "wasted" current through R_L in the ON state. Leakage current flows not only through all the collectors tied to R_L but also through all the transfer

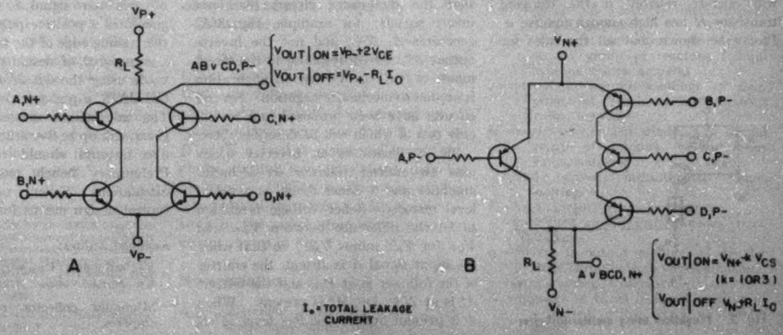
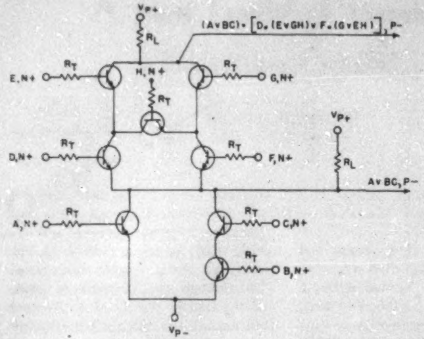


Fig. 3. Basic dual-range gates A—P-type gate of height 2 B—N-type gate of mixed height



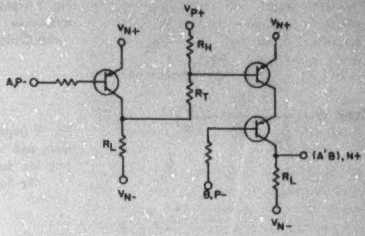


Fig. 4 (left). Combinational dual-range network

Fig. 6 (above). Negation in connection with AND-gate

resistors driven by the output; it is desirable that the output voltage remain more positive than V_{N+} in the OFF state. The wasted current is approximately $(V_{P+}-V_{P-})$ divided by R_L; large V_{P+}, within the limits of transistor body breakdown, and proportionately large R_L yield an optimum solution at the expense of increased power dissipation.

The circuits of Fig. 3 are limited in speed by transistor capacitances and hole storage. Switching speed can be increased by shunting each transfer resistor with a capacitor large enough to generate a transient voltage overshoot at the base during change of state. This is particularly effective in reducing hole (or electron) storage time by sweeping the holes out of the base region through reverse base voltage.

A more complicated *P*-type switching circuit is shown in Fig. 4. Two separate circuits have been coalesced to save equipment and two output leads are provided. In addition, the upper half of the circuit is a bridge circuit which will operate reliably if the bridging transistor *H* has high enough inverse α. It can be shown that all the rules for designing many-input many-output relay switching networks apply to dual-range network design as well.[8]

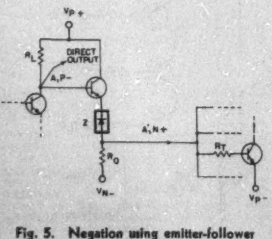


Fig. 5. Negation using emitter-follower

All transfer resistors have been indicated in Fig. 4 as having the same resistance R_T. A more efficient design might call for increased resistance for transistors closer to the output lead, since the bottom transistors carry not only the load current but also the base currents of all transistors above them. For large common-emitter current gain $(\beta \approx 50)$ use of many resistance values improves the circuit not more than 5%, based on the number of bases which can be driven by the output, and this is probably not enough to justify the added complexity. For low-gain transistors, $\beta \approx 10$, about 35% improvement results and the added complexity may be warranted.

As a rough guide it may be assumed that with ±10% resistance tolerance and ±5% supply voltage fluctuation, the number of loads which can be driven from a single output lies between 0.4 β and 0.5 β.

It has been shown in the foregoing that the dual-range circuit does not invert signals; for example, Fig. 3(A) generates *A*, *N*+, and not the inverse output *A'*, *N*+ in response to the direct input *A*, *P*−. A new configuration is needed to mechanize negation. Several circuits have been proposed and tested, only two of which will be described here.

Fig. 5 illustrates an inverter which uses an emitter follower as a buffer amplifier and a Zener diode as a signal level restorer. Zener voltage is chosen to be the difference between V_{P+} and V_{N+} (or V_{P-} minus V_{N-}), so that when the input signal *A* is absent, the emitter of the follower is at *P*+ and the output *A'* is at *N*+, i.e., *A'* is present. When *A* is present and at the *P*− level, *A'* is held absent at the *N*− level. The resistor R_Q provides enough quiescent current in the *N*− state to maintain Zener voltage across the diode. If *A'* is needed at only one point in the system, the emitter follower can be omitted from the inverter circuit. Note that the inverter connects a *P*-type circuit to another *P*-type circuit (or an *N*-type to an *N*-type).

In Fig. 6 negation is mechanized in connection with a simple switching circuit. This circuit is similar to the previous one without emitter follower but the Zener diode has been replaced by a resistor which provides the necessary voltage translation. This inverter also interconnects similar circuits, both *N*-type in this instance.

Pulse circuits and multivibrators, both monostable and astable, can be added to the dual-range family in the conventional manner through the use of coupling capacitors. Two differentiators of simple design are shown in Figs. 7(A) and 7(B). The former generates a negative pulse δA, *P*− in response to the leading edge of each input signal *A*, *P*−. The latter generates a positive pulse in response to the trailing edge of the same input signal.

A number of design calculations were made using the *GE 2N43* p-n-p and the *GE 2N78* n-p-n transistors as a basis. The calculations showed that five-high gates with up to five branches and driving five outputs should operate reliably. Preliminary bench tests checked the calculations. Salient numbers from a typical design are as follows.

Assumed values:

$V_{P+}=6$ volts; $V_{N+}=0$; $V_{P-}=-6$ volts; $V_{N-}=-12$ volts.

Minimum collector current = 2 milliamperes.

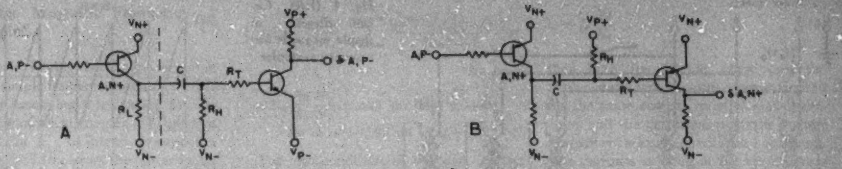


Fig. 7. Differentiator. A—For turn-on transition. B—For turn-off transition

Maximum ambient temperature = 45 C (degrees centigrade).

Tolerance on resistors = ±10%.

Tolerance on supply voltages = ±10%.

	GE 2N43, P-N-P		GE 2N78, N-P-N
Measured values			
Minimum β	20	...	20
ON collector voltage, volts	0.06–0.1	...	0.04–0.17
ON base voltage, volts	0.15–0.17	...	0.2 –0.3
Minimum OFF base voltage, volts	0.1	...	0.1
Maximum I_{co} at 45 C, μa	60	...	24
Design values for five-high five-wide network driving fine outputs, kilohms			
R_L	5.6	...	6.8
R_T	27	...	27

Conclusions

This paper has described a new dual-range family of transistor switching circuits with a number of properties which serve to advantage in switching systems. The component count is smaller than for conventional circuits patterned after vacuum-tube designs but it is greater than for DCTL systems.

The dual-range family offers a possibility of low stand-by power dissipation and should be capable of higher speed switching than the DCTL family. However, both p-n-p and n-p-n transistors are required in these dual-range switching circuits.

References

1. A Digital Computer for Use in an Operational Flight Trainer, W. H. Dunn, C. Eidert, P. V. Levonian. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., vol. EC-4, June 1955, pp. 55–63.

2. Transistor Circuitry for Digital Computers, C. L. Wanlass. *Ibid.*, vol. EC-4, Mar. 1955, pp. 11–16.

3. Electronics in Telephone Switching Systems, A. E. Joel. *Bell System Technical Journal*, New York, N. Y., vol. 35, Sept. 1956, pp. 991–1018.

4. Dial Telephone Switching Circuits. *Report No. 56-38*, Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa., section 3, 1956.

5. Surface-Barrier Transistor Switching Circuits, R. H. Beter, W. E. Bradley, R. B. Brown, M. Rubinoff. *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 4, 1955, pp. 139–45.

6. The Design of Switching Circuits (book), W. Keister, A. E. Ritchie, S. H. Washburn. D. Van Nostrand Company, Inc., Princeton, N. J., 1951.

Magnetic Frequency Multipliers and Their Rating

Part II—Frequency Sextuplers

WILLIAM McMURRAY
ASSOCIATE MEMBER AIEE

THIS PAPER presents an extension of the work reported in Part I.[1] The mode of operation and the rating of two frequency sextupler circuits are now analyzed by the method which was applied to frequency triplers in Part I. The circuits employ saturable reactors which are biased by means of direct current, and are very similar to the reactors used in the frequency triplers.

Simple Magnetic Frequency Sextupler

In the symmetrical magnetic frequency tripler circuit (Part I, Fig. 6), it is seen that the voltage developed across the choke L has a saw-toothed wave shape of sextuple frequency; see Part I, Fig. 9-(C). If a load is connected across the choke, sextuple frequency power may be obtained. Slight modification of the symmetrical frequency tripler circuit results in the simple frequency sextupler circuit shown in Fig. 1.

Nomenclature

e_{0A} = line-to-neutral voltage, phase A, instantaneous value, volts
E_m = line-to-neutral voltage, maximum value, volts (base of per-unit voltage)
e_c = voltage across choke, instantaneous value, volts
e_l = load voltage, instantaneous value, volts
E_l = load voltage, rms value, volts
e_r = interstage voltage, instantaneous value, volts
E_r = interstage voltage, rms value, volts
f = frequency of a-c supply, cps (cycles per second)
I_d = direct current bias, amperes (base of per-unit current)
I_d' = direct current bias on doubler stage, amperes
$i_{A,B,C}$ = currents in input windings of first group of reactors, instantaneous values, amperes
$i_{A,B,C}'$ = currents in input windings of second group of reactors, instantaneous values, amperes
i_0 = current in output windings of first group of reactors, instantaneous value, amperes
i_0' = current in output windings of second group of reactors, instantaneous value, amperes
i_l = load current, instantaneous value, amperes
I_l = load current, rms value, amperes
i_r = interstage current, instantaneous value, amperes
i_s = current in output windings of doubler stage, instantaneous value, amperes
n = order number of harmonic
N_t = number of turns in input winding of a reactor (prime refers to doubler stage)

Paper 57-91, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 17, 1956; made available for printing November 21, 1956.

William McMurray is with the General Electric Company, Schenectady, N. Y.

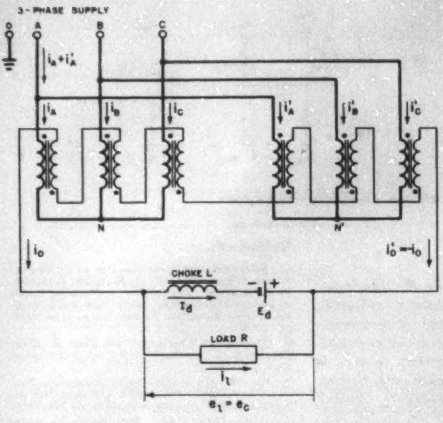


Fig. 1 (left). Circuit diagram of simple magnetic frequency sextupler

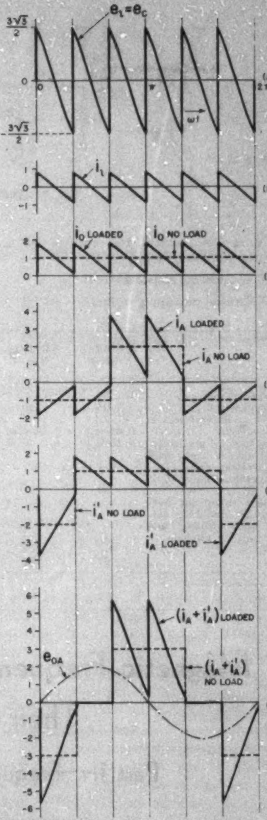


Fig. 2 (right). Wave forms in simple frequency sextupler

N_0 = number of turns in output winding of a reactor (prime refers to doubler stage)

P_l = load power, watts

t = time, seconds

$\omega = 2\pi f$

STEADY-STATE OPERATION

With the load R disconnected so that only the d-c bias current I_d flows in the output windings of the reactors, the voltages and currents existing in the simple frequency sextupler circuit are the same as for the no-load condition of the symmetrical frequency tripler; see Part I, Figs. 7, 8, 9(A) and (C), and 11(A). The load voltage e_l is the voltage e_c appearing across the choke L and may be expressed by the Fourier series as follows:

$$e_l = e_c = E_m \frac{N_0}{N_i} \frac{18\sqrt{3}}{\pi} \sum_{n=1}^{\infty} \frac{6n}{(6n)^2-1} \sin 6n\omega t$$

$$= E_m \frac{N_0}{N_i}(1.701 \sin 6\,\omega t + 0.833 \sin 12\,\omega t + \ldots) \quad (1)$$

For the case of resistive loading, the load current i_l, Fig. 2(B), has the same saw-toothed wave form as the load voltage e_l; see Fig. 2(A). The output windings carry the load current superposed upon the d-c bias, Fig. 2(C). The current in the output windings is reflected into the input windings and supply lines as shown in Fig. 2(D), (E), and (F), for phase A.

As is the case with the frequency triplers discussed in Part I, the present circuit exhibits a load current-limited mode of operation when the load resistance is less than a certain critical value. The critical load corresponds to the condition where the peak value of the load current equals the bias current. In the current-limited region, the load voltage drops rapidly, Fig. 3, as a result of the unsaturation of additional cores during parts of the cycle. Since this mode is not of great practical interest, it will not be discussed further in this paper.

RATING AND PERFORMANCE DATA

The ideal performance and rating factor of the simple frequency sextupler are calculated by a procedure similar to the example of Part I. For the case of unfiltered output into resistive load, one obtains the data presented in Table I. Tests were made upon an experimental circuit similar to the circuit of Fig. 1, except that a separate bias winding was used to prevent any direct current from flowing through the load. The extra winding, of course, increases the rating factor. The reactors were the same as were employed in the experimental frequency triplers of Part I. To approach the ideal conditions more closely, the choke was much larger than would normally be used. The characteristics shown in Fig. 3 were obtained with a supply of 430-volt, 3-phase, 60-cps, and a d-c bias of 5 amperes. For comparison, the ideal characteristic $E_l = 1.528$ per unit up to $I_l = 0.588$ per unit is also shown in Fig. 3. It is seen that there is considerable regulation due to high internal impedance in the actual circuit. In view of the poor wave shape, high rating factor, low power factor and poor regulation, the simple frequency sextupler is not suitable when large power output is required.

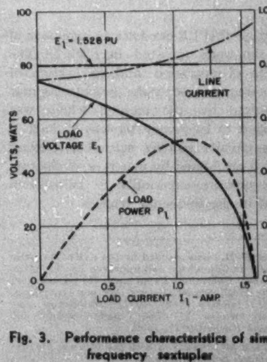


Fig. 3. Performance characteristics of simple frequency sextupler

Two-Stage Magnetic Frequency Sextupler

A circuit having better performance than the simple frequency sextupler is obtained by cascading a frequency doubler with a symmetrical frequency tripler, as shown in Fig. 4. An interstage capacitor *C* may be used to correct the power factor of the doubler stage.

STEADY-STATE OPERATION WITHOUT INTERSTAGE CAPACITOR

The frequency tripler stage operates as described in Part I, while the operation of the frequency doubler stage is essentially as described in reference 2, except that its supply voltage e_r, Fig. 5(A), is the output voltage of the tripler stage and has an almost rectangular wave shape. The load voltage e_l, as shown in Fig. 5(B) is the output voltage of the doubler stage, and also has a near-rectangular wave shape, which is given by the Fourier series:

$$e_l = E_m \frac{N_0}{N_t}\frac{N_0'}{N_t'}\frac{9}{\pi}\sum_{n=1}^{\infty}\frac{6n[2-\sqrt{3}(-1)^n]}{(6n)^2-1}\times \sin 6n\omega t$$

$$= E_m \frac{N_0}{N_s}\frac{N_0'}{N_t'}(1.833 \sin 6\omega t + 0.064\times \sin 12\,\omega t + 0.596 \sin 18\,\omega t + \ldots) \quad (2)$$

For the condition of no load on the doubler stage, the doubler acts as a nonlinear inductive load on the tripler. With resistive loading of the doubler stage, a load current i_l, Fig. 5(C) flows and the interstage current i_r, Fig. 5(E) acquires a component in phase with the voltage e_r. The doubler stage enters a current-limited mode of operation when the load resistance becomes less than a certain critical value corresponding to a peak value of load current equal to the d-c bias I_d' on the doubler. Similarly, the tripler stage cannot supply an interstage current having a peak value greater than twice the d-c bias I_d on the tripler. Thus, if

$$I_d' = I_d \frac{N_t'}{N_0'} \quad (3)$$

then both the doubler stage and the tripler stage will become current limited for the same critical value of load resistance, and the performance of the 2-stage frequency sextupler will be optimum. The currents existing in the tripler stage at no load and at full (critical) load are shown in Fig. 6 for the case when I_d' equals I_d, unity turns ratio being assumed. During the current limited mode of operation, in which the load resistance is less than the critical value, the circuit becomes a source of constant sextuple frequency current with rectangular wave shape of magnitude I_d', provided that equation 3 is satisfied.

RATING AND PERFORMANCE DATA

The rating factor and other ideal performance data for the 2-stage frequency sextupler circuit of Fig. 4 with purely

Table I. Rating Data for Ideal Magnetic Frequency Sextuplers

Quantity		Simple Sextupler, Fig. 1		Two-Stage Sextupler, Fig. 4 Tripler Stage		Two-Stage Sextupler, Fig. 4 Doubler Stage	
Load voltage	peak value	$\frac{3\sqrt{3}}{2}$	2.598	$\frac{3}{2}$	1.5	$\frac{3}{2}$	1.5
	average value	$\frac{18\sqrt{3}}{\pi}\left(1-\frac{\sqrt{3}}{2}\right)$	1.330	$\frac{9}{2\pi}$	1.432	$\frac{9}{2\pi}$	1.432
	effective rms value	$3\sqrt{\frac{3}{2}}\sqrt{1-\frac{3\sqrt{3}}{2\pi}}$	1.528	$\frac{3}{2\sqrt{2}}\sqrt{1+\frac{3\sqrt{3}}{2\pi}}$	1.434	$\frac{3}{2\sqrt{2}}\sqrt{1+\frac{3\sqrt{3}}{2\pi}}$	1.434
	maximum value of fundamental component	$\frac{108\sqrt{3}}{35\pi}$	1.701	$\frac{27\sqrt{3}}{8\pi}$	1.861	$\frac{54(2+\sqrt{3})}{35\pi}$	1.833
Voltage rating of reactor		$\frac{1}{\sqrt{2}}$	0.707	$\frac{1}{\sqrt{2}}$	0.707	$\frac{3}{8\sqrt{2}}$	0.265
Critical load resistance		$\frac{3\sqrt{3}}{2}$	2.598			$\frac{3}{2}$	1.5
Maximum load current, rms value		$\sqrt{2}\sqrt{1-\frac{3\sqrt{3}}{2\pi}}$	0.588			$\frac{1}{\sqrt{2}}\sqrt{1+\frac{3\sqrt{3}}{2\pi}}$	0.956
Maximum load power		$3\sqrt{3}\left(1-\frac{3\sqrt{3}}{2\pi}\right)$	0.899			$\frac{3}{4}\left(1+\frac{3\sqrt{3}}{2\pi}\right)$	1.370
Current in output winding, rms value	no load	1	1.0	$\frac{\sqrt{5}}{2}$	1.118	1	1.0
	maximum load	$\sqrt{3}\sqrt{1-\frac{\sqrt{3}}{\pi}}$	1.160	$\frac{1}{2\sqrt{2}}\sqrt{11+\frac{3\sqrt{3}}{2\pi}}$	1.216	$\sqrt{\frac{3}{2}}\sqrt{1+\frac{\sqrt{3}}{2\pi}}$	1.383
Current in input winding, rms value	no load	$\sqrt{2}$	1.414	$\sqrt{\frac{5}{2}}$	1.581	[illegible]	1.0
	maximum load	$\sqrt{6}\sqrt{1-\frac{\sqrt{3}}{\pi}}$	1.641	$\frac{1}{2}\sqrt{11+\frac{3\sqrt{3}}{2\pi}}$	1.720	$\sqrt{\frac{3}{2}}\sqrt{1+\frac{\sqrt{3}}{2\pi}}$	1.383
Line current, rms value	no load	$\sqrt{6}$	2.449	$\sqrt{\frac{13}{2}}$	2.550		
	maximum load	$3\sqrt{2}\sqrt{1-\frac{\sqrt{3}}{\pi}}$	2.842	$\frac{1}{2}\sqrt{27+\frac{3\sqrt{3}}{2\pi}}$	2.638		
Rating factor of one reactor		$\frac{(\sqrt{2}+1)\sqrt{1-\frac{\sqrt{3}}{\pi}}}{6\sqrt{2}\left(1-\frac{3\sqrt{3}}{2\pi}\right)}$	1.102	$\frac{(\sqrt{2}+1)\sqrt{11+\frac{3\sqrt{3}}{2\pi}}}{3(1+3\sqrt{3}/2\pi)}$	0.757	$\frac{\sqrt{3}\sqrt{1+\frac{\sqrt{3}}{2\pi}}}{2(1+3\sqrt{3}/2\pi)}$	0.268
Total rating factor of circuit			6.61		5.08		
Effective power factor of circuit			0.149		0.245		0.691

All values are per unit for unity turns ratio on the reactors. Bases of per-unit quantities: voltage, E_m (maximum value of line-to-neutral voltage, volts); current, I_d (d-c bias, amperes); resistance, E_m/I_d; power, $E_m\, I_d$.

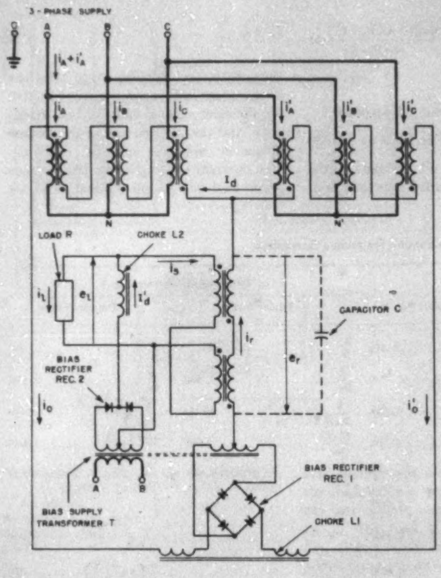


Fig. 4 (left). Circuit diagram of 2-stage magnetic frequency sextupler

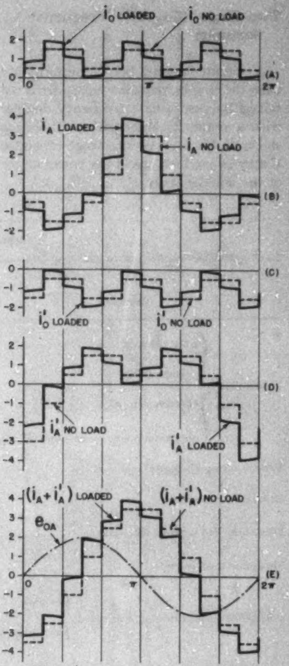


Fig. 6 (right). Wave forms in frequency tripler stage of 2-stage frequency sextupler

resistive load are given in Table I. Experimental characteristics of a typical circuit are shown in Fig. 7, which has been obtained from reference 3. Separate bias windings were employed on the reactors in the test circuit. The tripler stage is the same as was used to obtain the test data in Part I. The doubler stage consists of a pair of the following saturable reactors:

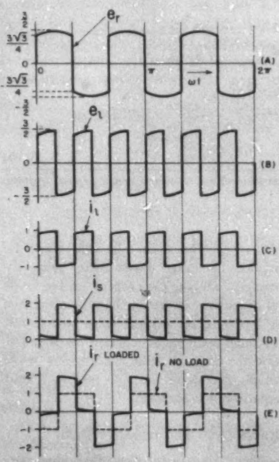


Fig. 5. Wave forms in frequency doubler stage of 2-stage frequency sextupler

Core. Arnold Engineering Company toroidal, tape-wound core Catalog no. *5468-D2* (0.002-inch Deltamax tape). Inside diameter 2.5 inches, outside diameter 3.5 inches, height 1.0 inch.

Windings. Input winding 120 turns 0.0571 inch wire; output winding 160 turns 0.0508 inch wire; bias winding 80 turns 0.0508 inch wire.

The data for Fig. 7 was obtained under the following conditions:

Supply voltage 450 volts, 3 phase, 60 cps.

Tripler stage d-c bias 5.0 amperes.

Doubler stage d-c bias 2.5 amperes.

Optimum performance is obtained when the d-c bias on the doubler stage is about 85% of the value given by equation 3. It is seen from Fig. 7 that the load voltage regulation of the 2-stage circuit is considerably better than the regulation of the simple frequency sextupler; reference may be made to Fig. 3.

The performance of the 2-stage sextupler may be improved by using an interstage capacitor *C* to correct the power factor of the doubler stage, enabling the rating of the tripler stage to be reduced. However, the capacitor distorts the interstage voltage e_r, and the load voltage e_l assumes a saw-toothed wave shape. Analysis of the mode of operation becomes difficult and will not be attempted here. Further information is given in reference 3.

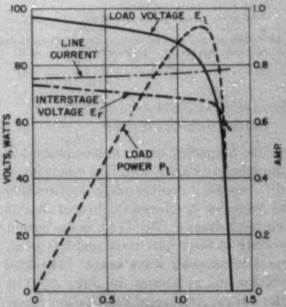


Fig. 7. Performance characteristics of 2-stage frequency sextupler

Discussion

The two magnetic frequency sextupler circuits described are presented as typical examples of the many possible circuits. In performance, the circuits presented are neither the best nor the worst. For example, one available circuit[4] has the same load voltage wave shape and same power factor as the 2-stage circuit, but has a rating factor of only 3.11.

The maximum load power has been assumed to occur with the critical value of load resistance, for the purpose of calculating the rating factor. In the frequency multiplier circuits which produce a saw-toothed load voltage, the actual maximum power output may occur at some point in the current limited region. However, there are considerable voltage regulation and wave form distortion at this point, and this is therefore assumed to be an undesirable operating condition in most applications. For applications such as induction heating, this condition may be acceptable and a somewhat lower rating factor is obtained.

In Figs. 2, 5, and 6, showing ideal voltage and current wave forms, the scales are marked in per-unit quantities and unity turns ratio on all reactors has been assumed.

References

1. Magnetic Frequency Multipliers and Their Rating, Part I—Frequency Triplers, William McMurray. *AIEE Transactions*, vol. 75, pt. I, Sept. 1956, pp. 384–90.

2. Magnetic Frequency Conversion, L. C. Harriott. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 9, 1953, pp. 78–87.

3. An Introduction to the Theory of Magnetic Frequency Multipliers Using Biased Magnetic Cores, D. W. Leiby. *M.S. Thesis*, Union College, Schenectady, N. Y., 1956.

4. An Extension of the Theory of Magnetic Frequency Multipliers, W. McMurray. *M.S. Thesis*, Union College, Schenectady, N. Y., 1956.

◆

Discussion

Paul P. Biringer (University of Toronto, Toronto, Ont., Canada): This is a welcome basic treatment of a family of frequency multipliers. I do not want to criticize the paper because I feel that it deserves only credit. It seems appropriate, however, to make a few additions for the benefit of the increasing number of people who intend to use static frequency multipliers.

The very high total rating factor of the discussed circuits can be considerably improved by using leading power factor loads instead of resistive ones. This leading power factor load will change the voltage-current characteristic of the circuit and may be adjusted so as to give a compounding effect. Both for triplers and doublers it is fairly easy to achieve an improvement in the rating factor with a value of 2 or more by this method.[1] Thus it is not necessary that a frequency sextupler should have five or six times the size of a transformer of equivalent rating.

A second remark is with regard to the indicated low power-factor of frequency sextuplers. The low power-factor is more of an obstacle than the high rating factor, mainly because it is not so simple to correct. The current drawn from the line will not be sinusoidal; thus power-factor-correcting capacitors will compensate only the fundamental frequency component of the line current. This fact indicates the need of special connections or equipment to achieve high input power factor.

REFERENCE

1. The Magnetic Frequency Doubler with Capacitive Load, Paul P. Biringer. *Research Report No. 7*, University of Toronto, Toronto, Ont., Canada, Sept. 1956.

William McMurray: The author wishes to thank Dr. Biringer, and agrees that the potentialities andlimitations of magnetic frequency multipliers should be clarified for the benefit of those interested in using them.

The effect of capacitors in the load circuit of a frequency multiplier may be divided into two cases, the boundary between which is indistinct. In the first case, a series capacitor may be used to compensate for the effective internal inductance of the multiplier, without affecting its basic mode of operation so long as the peak value of load current is less than the d-c bias. (Such a series capacitor is often used with high-frequency alternators.) For circuits which have a common winding for bias and output (such as the circuits discussed in the paper) the series capacitor also serves to block direct current from the load. In addition, the internal inductance and series capacitor act as a partial filter which improves the wave shape of the load voltage. Since the internal inductance of a frequency multiplier is generally quite high, especially when the core material departs appreciably from a square loop, the use of series capacitance is almost always beneficial and in some cases is essential for satisfactory performance. This is particularly true with poorer circuits such as the basic tripler and simple sextupler. As noted by Dr. Biringer, a marked improvement in the actual rating factor is often achieved. However, the actual rating factor still remains less than the ideal rating factor calculated in the paper, where the internal inductance is neglected. Academically, the filtering action may reduce the ideal rating factor (compare equations 20 and 28, Part I). To minimize the volt-ampere load on the frequency multiplier, the power factor of an inductive load should be corrected by means of a parallel capacitor. Similarly, if the load is too capacitive, such as that of fluorescent lamps with capacitor ballasts, power factor correction by means of parallel inductance may be required.

The second case of capacitive loading occurs when the current-limited mode of operation is entered. For example, if a capacitor is connected directly across the output terminals of a frequency multiplier, the sudden change in voltage due to "chopping" will attempt to send a high peak of current through the capacitor, but this will be prevented by the current-limiting action. Instead, the voltages developed by the frequency multiplier will be forced to change and an entirely different mode of operation, akin to what is termed "ferroresonance," will result. The magnitude and wave form of the output voltage, and of the input current, can show marked changes which depend upon the size and location of the load impedances. Bistable conditions are possible; i.e., two stable modes of operation may exist with the same values of load parameters. Other effects, such as subharmonic oscillations and self-biasing (an increase in the direct current) have been observed by the author in certain circuits. Analysis of this type of operation is difficult, and performance is best determined by experiment. Under certain conditions, some improvement in rating factor may be achieved. The change in output characteristics with load is an undesirable feature in many applications. However, if the load is constant, or if it can tolerate these changes, the ferroresonant mode of operation may be advantageous.

To summarize, the performance of a frequency multiplier can be improved by use of suitable capacitors, but caution is recommended when operation in the ferroresonant mode is contemplated.

The low-input power factor can be tolerated without correction in units of small size. In larger units, the better circuits having higher intrinsic power factor are chosen. The fundamental frequency component of line current can be corrected by capacitors in the normal way. The harmonic content can be greatly reduced by inserting inductors in series with the power factor correction capacitors to resonate at the harmonic frequency which is most prominent in the current. By this means, quite good power factor correction can be accomplished.

Measurements of Charge Imparted to Fine Particles by a Corona Discharge

GAYLORD W. PENNEY
FELLOW AIEE

ROBERT D. LYNCH
ASSOCIATE MEMBER AIEE

In electrostatic precipitation the diffusion mechanism of charging fine particles is extremely difficult to treat mathematically unless rather drastic simplifying assumptions are made. The size of these fine particles is comparable to the mean free path of the molecules of the gas. To treat this precisely one would have to include the effect of the field on the individual paths of the ions and consider the fact that the paths of the ions in the electric field are curves and not straight lines. If all of this is included in the differential equation it appears almost impossible to obtain any exact solution. White[1] has given an excellent review of the solutions which have been obtained. In order to obtain reasonable mathematical equations all investigators have neglected the effect of the main electric field and have treated the problem by kinetic theory methods. Of course, since the field was neglected in deriving the equations, the calculated charge is independent of the field. In the past the Millikan oil-drop method has been the usual method for measuring the charge and with this method it was difficult to obtain sufficient experimental information to determine the agreement with calculations.

The development of methods for making dioctylphthalate aerosols of uniform and controllable particle size provides a very useful tool for improving our knowledge of some basic processes in electrostatic precipitation. Since the particles are spherical, uniform in size, and adhere to collecting surfaces they can be used to check the ideal electrical performance of a precipitator. When some tests of this type were made on a 2-stage low-voltage precipitator, the efficiency obtained was higher than that predicted by the generally accepted theory. This and similar questions prompted an experimental study of the charging of fine particles.

Apparatus and Methods of Test

In this investigation a dioctylphthalate aerosol was generated which had particles substantially uniform in size. The particle size was determined by an optical instrument called the "Owl," which has been described by LaMer and Sinclair.[2] This instrument passes a parallel beam of white light through the aerosol, and for particles from 0.2- to 1.0-micron radius the particle size is determined by measuring the angle of reflection of red bands of light. From 0.1 to 0.2 micron in radius the size is determined by comparing the relative intensities of the components of vertical and horizontal plane polarized light reflected from the particles.

The aerosol was passed through the charging region and the resulting charge determined by one of two methods. In the "weight" method, the aerosol is drawn through a filter. The increase in weight of the filter gives the weight of the particles. The current collected is measured by a sensitive galvanometer so that the integral of current versus time gives the charge. Then by use of the size as determined by the owl, the charge per particle can be calculated. In the second method the mobility of the particles is determined from measurements of efficiency when the charged aerosol is passed through a special parallel plate precipitator. With the mobility and particle size known, the charge can be calculated from Stoke's law and Cunningham's correction.

The use of these two methods is believed to provide a good check because the methods are quite different. For example, in calculating the charge per particle by the weight method the charge varies as the cube of the diameter whereas in the mobility method it varies approximately as the first power of the diameter. Then if one questions the size as determined by the Owl, the charge, as determined by the weight method, can be equated to the charge by the mobility method leaving size as the unknown. This equation can then be solved to determine the size of particle independently of the owl measurement.

The aerosol generator was of the type developed by the Naval Research Laboratory[3] which in turn was patterned after one built by the Massachusetts Institute of Technology CWS Development Laboratory. In this a hot air stream is saturated with dioctylphthalate (DOP) vapor. This hot air stream is suddenly mixed with a larger stream of cool air. The DOP vapor condenses which forms fine particles when the two air streams mix. The first generator built did not give a uniform particle size. This was in agreement with the general reputation of this type of generator which seemed to be that it could generate large volumes of aerosol but that the particle size would be nonuniform.

Rebuilding the smoke generator to insulate each air stream carefully and to hold temperatures constant up to the point of mixing gave a marked improvement. The particle size was now substantially uniform, and by varying the temperatures of the air streams, some control of particle size could be obtained. The uniformity of particle size was checked by growing the particles with toluene vapor[4] to a size between 0.2 and 0.6 micron in radius. Then if this aerosol was placed in the Owl and sharp red bands observed, it was assumed that the particle size was substantially uniform.

In the weight method a filter is needed which can efficiently remove fine particles and whose weight does not change appreciably with reasonable changes in humidity. For this purpose Fiberglas filter web type *MSA #1106-13* made by Mine Safety Appliance Company was used. This filter was found to remove particles effectively and its weight was reasonably independent of relative humidity. However the filter did absorb a measurable weight of DOP vapor. It was therefore necessary to remove the particles from the aerosol and pass the remaining vapor through the filter until its weight became constant before making a test. The current collected by the filter was measured by means of grounding the filter through a galvanometer

---

Paper 57-89, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted September 26, 1956; made available for printing January 9, 1957.

Gaylord W. Penney is with Carnegie Institute of Technology, Pittsburgh, Pa., and Robert D. Lynch is with Westinghouse Electric Corporation, Youngwood, Pa.

This investigation has been supported by the Aerotec Corporation. The authors wish to acknowledge the suggestions and encouragement of E. R. H. Pegg of the Aerotec Corporation and Sidney Orem, Jr., of the Thermix Corporation.

This work was done in part by R. D. Lynch in partial fulfillment of the requirements for the degree of doctor of philosophy at Carnegie Institute of Technology. The authors also wish to acknowledge the work of Arthur H. Hurst in conducting some of the tests.

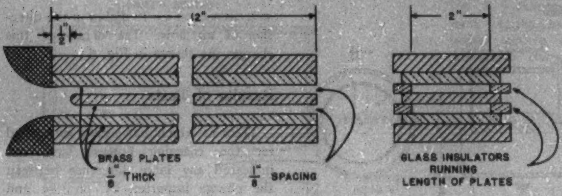


Fig. 1. Parallel plates used for measuring particle mobility

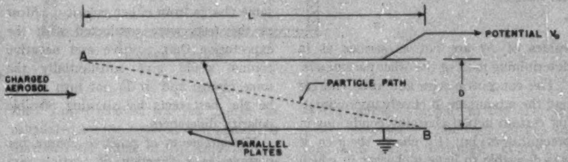


Fig. 2. Path of a charged particle entering at A for a potential (V_0) applied between the plates

having a sensitivity of 10^{-10} amperes per division.

The weight method was difficult to use because considerable time was required to collect a weighable amount of DOP. During this time all conditions of test must be held constant. Since it is difficult to maintain constant conditions in the smoke generator for long periods, many tests would be ruined by changes during a test.

In the mobility method, the charged aerosol was passed between parallel plates with a controllable potential between the plates. A brass plate was supported by glass insulating strips between two outside grounded plates. This forms two smooth rectangular air passages with a controllable potential between the parallel sides of each passage. This is a type of parallel plate precipitator in which turbulence has been minimized and in which the electric field is uniform. The details of the structure are shown in Fig. 1.

Since the electric field is uniform in this precipitating region, a charged particle experiences a constant electrostatic force perpendicular to the direction of gas flow. In an ideal situation with no turbulence and a uniform gas velocity this would give a constant component of velocity perpendicular to the axis of the duct and a constant component of velocity equal to the gas velocity parallel to the axis of the duct. This ideal path is shown by the dotted line of Fig. 2. This simplified analysis assumes uniform velocity, no turbulence, uniform particle size and charge. For these conditions it is obvious that efficiency is proportional to potential between the plates. Actually a uniform velocity cannot be attained over any appreciable length, but it can be shown that a variation in velocity across the duct does not disturb the linear relation between efficiency and potential. In the tests this linear relation could be realized up to 80 or 90% efficiency. A typical curve is shown in Fig. 3. Some slight turbulence and other disturbing influences were always present and decreased the efficiency in the region above 80%. The straight-line portion of the plot was used to compute mobility. Extension of this straight line to 100% efficiency gives the voltage required to move a particle across the width of the duct in the time required for air to traverse the length of the duct at its average velocity. (Results obtained by this method seem to agree with the weight method within the accuracy to be expected of such methods as illustrated by the points on the curve of Fig. 8.) Results could be obtained much more rapidly by the mobility method than by the weight method and so most curves were taken by the mobility method.

In the first weight tests a short tube connected the filter chamber to the charging chamber and the charge obtained by the weight method was materially higher than by the mobility method. Various observations were in agreement with the hypothesis that in addition to the uniform-size aerosol particles, there were a large number of much smaller particles probably in the range of 0.01 micron in diameter. Daniel and Bracket[5] have described tests indicating the existence of such particles in many situations. Such particles, having a diameter of the order of 0.01 micron, would not have sufficient light-scattering power to give a reading on the Penetrometer and would have little weight but would be able to carry a measurable charge. The testing procedure was changed to give a greater length of hose between the charging region and the filter chamber. This allowed time for the very small, high-mobility particles to diffuse to the wall. Under this condition reasonable agreement between weight and mobility methods was obtained.

Most of the tests used a concentric wire and tube corona discharge for charging the particles. A tungsten wire, 0.003 inch in diameter, was mounted on the axis of a tube with an inside diameter of 1.93 inches. The detail dimensions are shown in Fig. 4. The exact length of the corona is difficult to specify because of

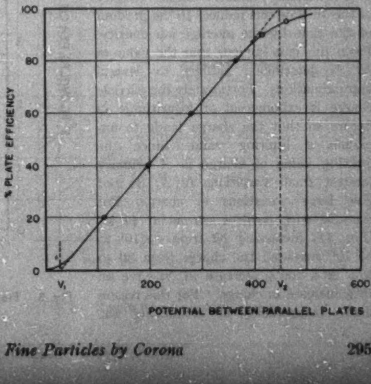


Fig. 3. Sample curve of the percentage of aerosol precipitated between plates versus the applied potential

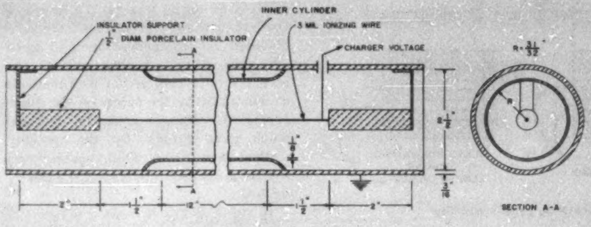


Fig. 4. Cylindrical charger

the end effect. In computing the time of charging, the effective length was assumed to be 13 inches.

Potential gradient as a function of radius for various wire potentials is plotted in Fig. 5. It is evident that for this range of potentials the gradient is reasonably constant except for the inner 25% of the radius. Because of the cylindrical geometry this inner 25% of the radius comprises only 6.25% of the volume. Experimental evidence indicates that the smoke density is greatly reduced in the inner high-field-strength region and so it seems reasonable to neglect this region in the charging process.

The ion density (N) in this cylindrical corona discharge is a function of radius. Visual observations of the corona using a small stream of smoke show that a strong turbulence or "electric wind" exists. Since it appears that each particle moves about radially during the charging period, it seems reasonable to use the average ion density in the outer 93% of the volume as the effective ion density.

All of the theories proposed for the charging process indicate that the charging rate is proportional to the ion density in the region surrounding the particle. For this reason one variable has been assumed to be $\int N dt$. If one assumes a substantially uniform linear velocity of the air flow this reduces to the product of the time by the average ion density. From this it is obvious that the value of Nt is determined subject to several approximations. Fortunately the particle charge is exponential or logarithmic in nature so that the charge tends to approach a limiting value. Since the limiting value of charge is of primary interest small variations in Nt do not give large variations in charge. For example with positive corona at 11 kv (Fig. 11) increasing Nt from 4×10^7 to 8×10^7 increased the charge from 30 to 35.5 or a 100% increase in Nt gave an 18% increase in charge. For this reason slight inaccuracies in computing the values of Nt are not as serious as in determining most of the other parameters.

The concentric wire and tube charger had the advantage of closely approximating certain actual charging conditions in precipitators but to check theories it was desirable to have a charger in which both field strength and ion density could be substantially uniform over the charging region. For this purpose a charger of the type developed by G. W. Hewitt was constructed. This uses the field between parallel grids or screens for charging particles. A relatively low ion density is required in order to have a uniform electric field. This requires a long charging time as compared to the concentric wire and tube to obtain a given value of Nt. A square wave of potential between the parallel plates was used to avoid precipitating the particles during the long charging time. Ions were supplied from both sides of the charging region so that both halves of each charging cycle were symmetrical. The charging region was 1 by $2^3/_4$ inches in cross section by 12 inches in the direction of air flow. The details of this charger are shown in Fig. 6.

Results of Tests

Theories which have been developed for the charging process have not indicated any large difference between the charge imparted by positive and negative corona provided voltage and Nt are held constant. Some preliminary tests at low voltage indicated about the same charge from either polarity. Most of the tests were conducted with the expectation that positive and negative corona would give substantially the same charge and so do not turn out to be the best tests for showing possible polarity differences.

Measurements of positive charge per particle as a function of particle size are shown in Fig. 7. In this case the charge per particle was measured by the filter paper, or weight method, and by the mobility method. This and other tests seem to indicate that both methods give the same charge within the accuracy of this type of measurement. Since the mobility test required much less time, only the mobility method was used in the remaining series of tests.

Tests were made with positive corona in the cylindrical charger in which the particle size and air flow were held constant and the wire potential varied. Since the time of charging was held constant, the value of Nt increased with wire potential or potential gradient in the charging space. The results of this series of tests are plotted in Fig. 8.

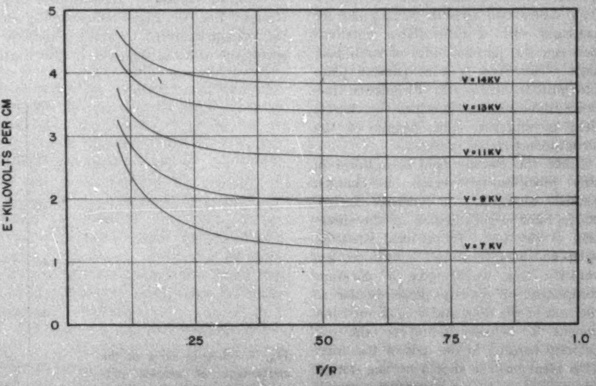


Fig. 5. Field strength in the cylindrical charger versus the ratio of the radius (r) to the radius of the cylinder (R) for various applied potentials

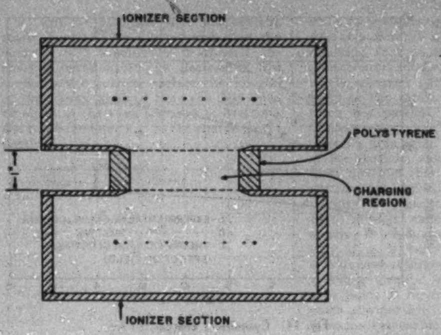


Fig. 6. Arrangement of the two ionizer sections of the screen charger

Also, one curve shows the charge as calculated by the conventional equations in which the effect of the field is neglected. It appears that, if the curve of measured charge and the curve for calculated charge could be extrapolated to zero field, they would both give the same value of charge. However, at a field of 4,000 volts per cm (centimeter) and for $Nt=10^8$ the measured charge is substantially twice the calculated charge. These curves show rather clearly the importance of the field in the charging process.

From a theoretical viewpoint the screen charger has the advantage that at any given instant both the ion density and field strength are substantially constant over the charging region which cannot be true in the cylindrical charger. Fig. 9 shows the results of a series of measurements using the screen charger with negative corona. In these tests the air flow was held constant and the voltages on the fine wires relative to their respective screens held constant. Under these conditions, the value of Nt decreases as the voltage gradient in the charging space increases. Again, the curves for calculated and measured charge, if extrapolated, would appear to intersect at zero voltage gradient. As in the case of the cylindrical charger at the highest voltage gradient, the measured charge is substantially twice the calculated value. In general the tests in the cylindrical charger were conducted at higher values of Nt than the tests in the screen charger and so it is difficult to compare results. However, at 1,600 volts per cm the value of Nt was 3.3×10^7 in both tests. For this condition both series of tests give about the same charge even though one is positive and the other negative.

Some tests which used negative corona near sparkover voltage of the cylindrical charger gave unexpected results. At low voltage the charge due to negative corona was only a little greater than that given by positive corona, which agreed with expectations. However, as the voltage on the wire increased the charge imparted by negative corona increased much more rapidly than that due to positive corona.

Curves of charge as a function of potential of the wire are given in Fig. 10. At 7 kv on the wire the curve drawn for the negative charge gives a value 12% greater than that for positive wire but as the voltage increases the curves rapidly diverge. At 13 kv the negative charge was 50% higher than positive. This was still not up to the sparkover region. It seems probable that if the voltage was increased until occasional sparkover occurred the difference would be still greater.

In another series of tests the potential on the wire was held constant and the rate of air flow varied. In this way Nt could be changed without changing other conditions. The results of these tests are shown in Fig. 11. The negative charge shows a rapid increase with Nt even for Nt above 10^8. It is suspected that the high charge given by negative corona is due to peculiar conditions near the breakdown region, primarily the presence of a very small percentage of free electrons.

At the lowest voltage gradients all tests give nearly the same charge provided the value of Nt is the same. At high-voltage gradients negative corona in the screen charger seems to give a charge approximately the same as positive corona in the cylindrical charger but much lower than negative corona in the cylindrical charger. This would be consistent with the hypothesis that the high charge was due to free electrons since in the screen charger the voltage beween the screens was far below breakdown. Conditions were therefore favorable for attachment of electrons to oxygen molecules before entering the charging region.

The high charge given by negative corona in this small cylindrical charger seems well established. But until measurements can be made with other sizes of

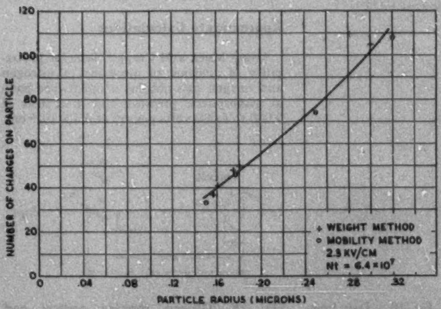


Fig. 7. Charge versus particle radius (cylindrical charger)

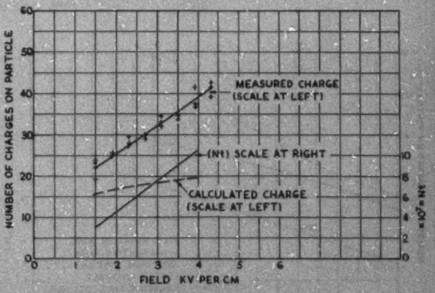


Fig. 8. Results from cylindrical charger

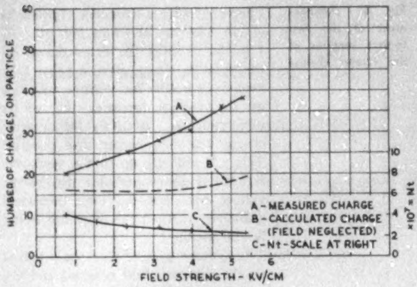


Fig. 9. Results from screen charger

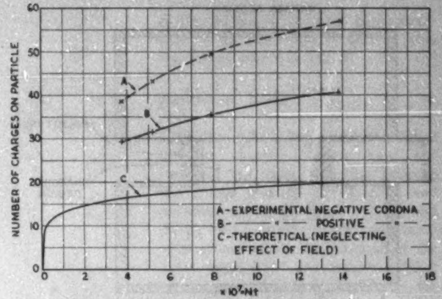


Fig. 11. Cylindrical charger, 11 kv

cylinders no general conclusions can be drawn. If the high charge is really due to a few free electrons then conditions in a larger cylinder with a longer time for attachment might be quite different. Some approximate calculations of charging rate for free electrons are given in the appendix. While these calculations are based on very rough approximations, they do seem to indicate that in the case of electrons the free path is sufficiently long, so that an electron has an appreciable probability of receiving enough energy from the field in a single free path to drive the electron to the particle against the repulsive field of the charge even with $n=40$ on a particle of 0.15-micron radius.

Conclusions

Experimental measurement of charge on particles of the order of 0.15 micron in radius, if extrapolated to zero field strength, would appear to agree satisfactorily with the values given by equations derived neglecting the effect of the main field on the diffusion charging process. However, at the fields used in commercial precipitators the effect of the field is not negligible, particularly at high values of Nt. With positive corona at the highest values of field strength and Nt tested the measured charge was about twice the value calculated neglecting the effect of the field.

With negative corona and the fine wire and small cylinder used for charging in these tests, a value of charge was measured which was 50% greater than the charge measured with positive corona at the same values of voltage and Nt. This high value of charge obtained with negative corona at higher values of gradient and Nt should be checked for larger wires, larger cylinders, and other electrode configurations. The authors suggest that a very small percentage of free electrons in the charging region may be present with negative corona near sparkover and that these free electrons could be largely responsible for these high values of charge obtained with negative corona in these tests.

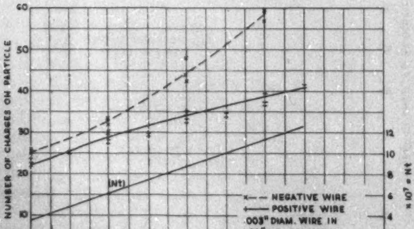


Fig. 10. Tests from cylindrical charger

Appendix. Charging Rate for Free Electrons

If any free electrons exist in the charging region they will have a mean free path which is long as compared to that of an ion. Accordingly it is desirable to examine the probability that an electron can receive sufficient energy from the field in a single free path to enable it to arrive at the particle against the repelling force of the charge on the particle.

Nomenclature

a = radius of particle, cm
e = elementary electronic charge: 4.8×10^{-10} statcoulomb
n = number of elementary charges on particle
t = time of exposure of particle to charging conditions, seconds
E = field strength in charging region, statvolts per cm
N = ion concentration in charging region, ions per cm^3
N_e = number of free electrons per cm^3
A_1 = projected area of the particle, cm^2
v = kinetic theory velocity, rms
x = distance from center of particle to the electron (D and E are assumed to be in the same direction)

Approximate Calculations

Using Meek and Craggs[8] data for probability of collision of electrons in nitrogen and oxygen one obtains a mean free path at atmospheric pressure and in a field of 3,000 to 5,000 volts per cm of 4.1×10^{-5} cm.

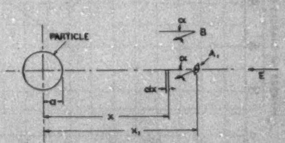


Fig. 12. Diagram showing particle and electron

It will be assumed that an electron starts from rest at a point such that the electric field accelerates it toward the charged particle. If the electron is just to reach the particle surface the integral of the energy received from the field must be equal to the work required to overcome the repulsive force due to the charge on the particle.

$$\int_{x_1}^{a} -Eedx - \int_{x_1}^{a} -\frac{ne^2}{x^2}\,dx = 0$$

$$E(a-x_1) - ne\left(\frac{1}{a}-\frac{1}{x_1}\right) = 0$$

If

$a = 1.5\times10^{-5}$ cm
$n = 40$
$E = 10$

Then

$x_1 = 12.8 = 10^{-5}$ cm

The electron must travel from x_1 to a or 11.3×10^{-5} cm. Assuming a Maxwellian distribution, the probability of a free path of 11.3×10^{-5} cm with a mean free path of 4.1×10^{-5} is

$$\frac{N_{x_1}}{N_e} = e^{-\frac{11.3\times10^{-5}}{4.1\times10^{-5}}} = e^{-2.76} = 0.0635$$

Thus one electron in 16 if traveling in the direction of the field will have a sufficient length of free path to receive the necessary energy to overcome the repulsive force due to 40 electronic charges on a particle of 0.15-micron radius.

The problem is to find the number of electrons having such a free path and terminating on the particle in unit time. To solve this problem one might find the number of collisions in any differential volume and the probability of moving in the direction of the particle with a sufficiently long free path to acquire enough energy from the field to arrive at the particle. The integration of this over the volume around the particle is extremely complicated mathematically. This solution would also require that the distribution of electrons in the region around the particle be known. Since this calculation cannot yet be made, an attempt will be made to evaluate the order of magnitude of the effect by considering electrons crossing an area at a distance x_1 from the particle and assuming that at this point the density of electrons is not disturbed by the proximity of the charged particle. To attempt to estimate the rate at which electrons can reach the particle, consider the number of electrons at point A_1 of Fig. 12 which cross an area equal to the projected area of the particle, assuming that they start from rest and that the length of path is not changed by the field. While the initial velocity is neglected here, the angle of that motion cannot be. An electron starting from A_1 can arrive at the particle only if α is very minute. However, if α is somewhat larger there will be a corresponding electron with the same angle α at a point B such that this electron can reach the particle. If α is not too large it seems reasonable to integrate over the angle α and over an area A_1 equal to the projected area of the particle and assume that these electrons which have a path X_1 or longer can reach the particle. The angle over which this approximation can be made is arbitrarily assumed to be from 0 to 30 degrees. The number of electrons per unit area per second with an angle α between 0 and 30 degrees is then the following

$$\int_0^{\pi/6}\left(\frac{N_e\mathcal{E}}{4\pi}\right)(2\pi\sin\alpha d\alpha) = 0.0665N_e\mathcal{E}$$

The rate at which electrons arrive at the particle is 0.0665 $N_e\mathcal{E}$ multiplied by the projected area of the particle and by the fraction which have a sufficiently long free path to overcome the repulsive field of the particle. If the number of ions is 10^8 per cm^3 and 1% are free electrons, then $N_e = 10^6$. Kennard[7] gives the thermal velocity of electrons at 15 degrees centigrade as 10.5×10^6 cm per second.

Then

$$\frac{\Delta n}{\Delta t} = 0.0665\times10^6\times10.5\times10^6\times \pi(1.5\times10^{-5})^2\times0.0635 = 31.3$$

This will be compared with the curves of Fig. 11 which correspond to a gradient of about 3,000 volts per cm. At 40 charges per particle $Nt = 4.2\times10^7$. If N is 10^8 then the time would be 0.42 second and 0.1 second later the charge would be 43.3 charges or a rate of charging of 33.0 charges per second. The particular number 31.3 is not significant since it is based on a pure assumption of one electron to 100 ions. In the calculation the thermal velocity of the electrons was neglected. At these gradients the temperature of the electrons is high so the initial velocity is not negligible and will give a marked increase in the number arriving. The paths of the electrons are curved and not straight which tends to decrease the effective area of the particle. Also, it is assumed that all electrons crossing the area A_1 start from rest whereas in the real situation those which cross already have used some of their path and will be less likely to arrive at the particle but other electrons will have paths starting inside of A_1 and so will be more probable to arrive. Considering the nature of the approximations the calculation given cannot be considered as any real evaluation of the charging rate but only as indicating that if free electrons are present they may account for an appreciable part of the charging rate at higher values of Nt.

(*Discussion on page 306*)

References

1. PARTICLE CHARGING IN ELECTROSTATIC PRECIPITATION, H. J. White. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1186–91.

2. PORTABLE OPTICAL INSTRUMENT FOR THE MEASUREMENT OF THE PARTICLE SIZE IN SMOKES, THE "OWL"; AN IMPROVED HOMOGENEOUS AEROSOL GENERATOR, V. K. La Mer, D. Sinclair. *Report 1668*, Office of Scientific Research and Development, National Defense Research Committee, Washington, D. C., Aug. 3, 1943.

3. DEVELOPMENT OF SMOKE PENETRATION METERS, H. W. Knudson, L. White. *Report no. P-2642*, Naval Research Laboratory, Washington, D. C., Sept. 14, 1945.

4. REPORT NO. NYO-514, V. K. La Mer. *Contract no. A.T.(30-1)-651*, Atomic Energy Commission, Washington, D. C., June 30, 1952.

5. AN ELECTRICAL METHOD FOR INVESTIGATING THE NATURE AND BEHAVIOR OF SMALL AIRBORNE CHARGED PARTICLES, J. H. Daniel, F. S. Brackett. *Journal of Applied Physics*, New York, N. Y., vol. 22, no. 5, May 1951, pp. 542–54.

6. ELECTRICAL BREAKDOWN OF GASES (book), J. M. Meek, J. D. Craggs. Clarendon Press, Oxford, England, 1953, p. 7.

7. KINETIC THEORY OF GASES (book), E. H. Kennard. McGraw-Hill Book Company, Inc., New York, N. Y., first edition, 1938 p. 26.

The Charging of Small Particles for Electrostatic Precipitation

G. W. HEWITT
MEMBER AIEE

CHARGING CONDITIONS and particle size are important factors in electrostatic gas cleaning. Particle mobility, the rate at which the particle moves in an electric field of unit strength, is a function of the charge and size of the particle. The minimum mobility and the particle size at which it occurs, are functions of the charging conditions.

Charging generally is accomplished by exposing the particles to a stream of ions of one sign in the gas. Three important conditions which affect the charging of a particle are electric field strength, ion concentration, and time of exposure to the ions. The acquired charge depends also upon physical properties of the particle, particularly its size.

Electrostatic precipitation now is used extensively in air-conditioning applications for the purpose of greatly reducing the blackening of walls, furnishings, and other exposed surfaces, by the atmosphere.[1] Fractional-micron particles are responsible for a large part of this blackening. A study of city atmosphere made at this laboratory using an electron microscope, indicates that about half the blackening is attributable to particles between 0.1 and 0.6 micron in radius. This is an important size range in the cleaning of the atmosphere.

Existing equations do not permit accurate calculation of the charging of these small particles, and further than this, there have been very few experimental data which have been published, particularly for practical charging conditions. The purpose of this paper is to present experimental information on the charging of fractional-micron particles in air, as a function of particle size, field strength, ion concentration, and charging time.

Experimental Methods

The isolation of particle size effects has been one of the difficulties in particle charging studies in the past. The effects of particle size in the fractional-micron range were difficult to determine because it was necessary to make measurements on individual particles. The study of clouds of particles is simpler and less tedious, and it automatically provides an average result for a large number of particles. For studies on clouds of particles, it is required that the size be nearly uniform and controllable over the desired range.

AEROSOL GENERATION AND MEASUREMENT

Cloud-type measurements were made possible by the use of an aerosol generator and light-scattering instruments developed at Columbia University during World War II by La Mer and Sinclair.[2-4] These instruments are described in Appendix I. The generator produces particles of controllable and very uniform size. The particles are tiny liquid spheres. The liquid used in these tests was diocytl phthalate, which has a very low vapor pressure at room temperature, and provides a very stable aerosol. The gas in which the particles were suspended was a mixture of nitrogen, oxygen, and water vapor in the proportion in which they occur in air at 20% relative humidity.

The degree of particle size uniformity produced by this type of generator is given by La Mer. Generally, more than 70% of the particles are within ±10% of the average radius, and a negligible percentage differs from the average by more than ±20%.

The size and concentration of the aerosol particles were measured by light-scattering methods developed at Columbia University, and described in Appendix I. The instrument used to determine particle size, code-named the "Owl," provides simple and rapid size measurements of uniform aerosols.

In the aerosol concentration meter, light scattered by the particles is focused upon a photomultiplier tube. The response of this instrument is linear with particle concentration over a wide range, the particle size being constant. In most of the tests only the relative concentrations, at constant particle size, were required. In some cases, the absolute concentration in terms of liquid weight per unit volume of air was required. For this purpose, the instrument was calibrated experimentally with respect to particle size.

PARTICLE CHARGER

The charging device was designed so as to provide independent control of the three important charging variables, field strength E, ion concentration N, and charging time t. This device is shown in Fig. 1. The charging region C is bounded on opposite sides by electrodes A and S between which a voltage is applied to obtain the desired field strength. Electrode S is a fine screen of large percentage open area. Ions of one sign (positive in the case of the polarity shown), produced by corona at coaxial wire W, are caused to flow through the screen and across the charging region. The aerosol flowing axially between A and S is subjected to these ions and to the voltage gradient in this region, for a time t which is determined by the aerosol velocity and the axial length D of the charging region. The cylindrical shutter H, slidable inside cylinder B, is used to adjust D. The ion concentration in region C is controlled by adjusting the voltage between the coaxial wire and the cylinder B.

As soon as the aerosol particles pick up some of the ions, they experience an electrostatic force moving them in the direction of the field. For many charging conditions, this would cause the loss of a large percentage of the particles by precipitation on electrode A if the voltage between S and A were continuous. To reduce loss by precipitation greatly, an alternating voltage is used between S and A, so as to limit the transverse motion of the particles to small amplitudes. During the reverse half-cycle, the direction of the field in C prevents ions from passing through the screen, so no charging occurs. A square wave of voltage is desirable in order that E and N remain constant during each charging period.

Variations in E and N with radial position in region C were less than 10% of the average values in most of the tests. Further information on this and other details of the charger are given in Appendix III.

MOBILITY ANALYZER

The mobilities of the charged particles were determined by measuring their velocities in an electric field. Fig. 2

Paper 57-90, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21-25, 1957. Manuscript submitted August 28, 1956; made available for printing February 6, 1957.

G. W. HEWITT is with the Westinghouse Electric Corporation, Pittsburgh, Pa.

The author acknowledges the assistance of Mrs. G. Lachmanek and M. C. Werst in carrying out this project, and the advice of Dr. A. V. Phelps, all of the Westinghouse laboratories; also, the advice of Professor G. W. Penney, Carnegie Institute of Technology.

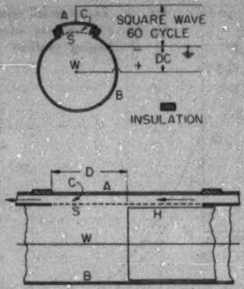


Fig. 1. Particle charger. Charging region is shown at C between electrode A and screen S. Ions are furnished by corona at wire W in cylinder B. Length D of charging region is adjustable by positioning of shutter H

shows the analyzer and the arrangement of apparatus. Clean air flows axially in a laminar stream between two coaxial cylinders which are held at different potentials. The charged aerosol enters the analyzer through a narrow annular passage *A* at the circumference of the outer cylinder. Downstream from this, a narrow circumferential sampling groove is provided in the wall of the inner cylinder at *B*. This sampling groove connects to an exit tube of small diameter inside the inner cylinder, leading out of the analyzer. The inner cylinder is movable axially for adjustment of the distance *L*. The mobility of the particles entering the sampling groove at *B*, is a function of the analyzer dimensions, the distance *L*, the air-flow rate, and the voltage *V* between cylinders, as given by equation 11. It is independent of the radial distribution of air velocity between the cylinders. This calculation and other information on the analyzer are given in Appendix IV.

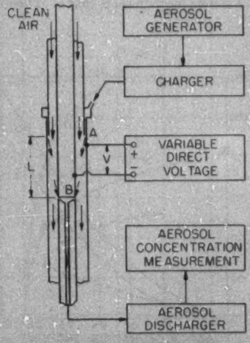


Fig. 2. Coaxial cylinder mobility analyzer. Charged aerosol enters at A. Sampling groove is at B in inner cylinder, which is adjustable axially

Upon leaving the analyzer, the aerosol sample is passed through a chamber in which it is exposed to alpha radiation. The purpose of this is to bring the net charge on the aerosol to near zero in order to minimize the loss of particles by space-charge precipitation, particularly in the light-scattering chamber of the concentration meter, through which the aerosol then flows.

For any given aerosol entering at *A*, the mobility test is made by holding the air flow and distance *L* constant at suitable values and recording the concentration of the sample, while varying voltage *V* in steps. From these data, a curve showing relative particle concentration is plotted against mobility; Fig. 3 is a typical curve. The average mobility is taken as the average of the mobilities at half the peak ordinate, as indicated by the broken lines.

The actual spread in particle mobility, caused by statistical variations in charging and a small variation in particle size, is smaller than the spread of the corresponding analyzer curve, because some broadening is caused by conditions in the analyzer. This is discussed in Appendix IV.

The relation between mobility and particle charge can be determined from mobility measurements on particles of the same size carrying very few charges. One way of charging particles very weakly is to subject the aerosol to ionizing radiation. This provides both positive and negative gas ions, so some particles become weakly charged negatively and some weakly charged positively. A preferable method is to use ions of only one sign in order to minimize agglomeration of the particles. This is done easily with the charger shown in Fig. 1 by using very small values of *E*, *N*, and *t*.

To measure the mobilities of these particles, low air velocity and large distance *L* in the analyzer are required. A typical test result is shown in Fig. 4. Note that the mobilities corresponding to the three peaks, are related in a 1-to-2-to-3 proportion. Since electric charge exists in discrete quantities, it is evident that the three peaks correspond to particles carrying one, two, and three elementary charges. This provides the linear relation between charge and mobility, which can be used to calculate charge from mobility measurements on particles of the same size, for various charging conditions. For example, since Figs. 3 and 4 are for the same particle size, the average particle charge for the test shown in Fig. 3 is 21.

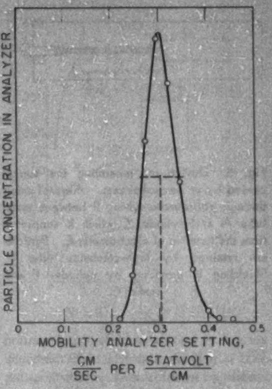


Fig. 3. Typical analyzer response curve. Data obtained on particles of 0.13-micron radius charged at $E=12$, $Nt=10^7$

CHARGE-MASS MEASUREMENTS

Another method of determining particle charge, independent of mobility measurements, is that of measuring the charge per unit mass of liquid in the droplets. From this the particle charge can be calculated readily for a uniform

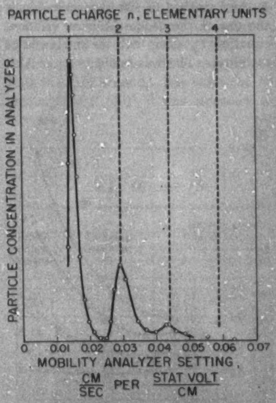


Fig. 4. Analyzer response curve for very weakly charged particles, 0.13-micron radius. Equally spaced peaks indicate the particle charge

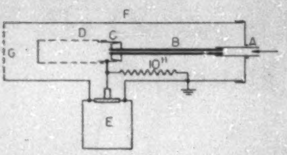


Fig. 5. Device for measuring the current carried by an aerosol stream. Aerosol enters through polystyrene tubing B between metal tube A and cylinder C which is supported from the terminal of electrometer E. Particles are retained by high-efficiency filter D. Shielding is provided by cylinder F and screen G

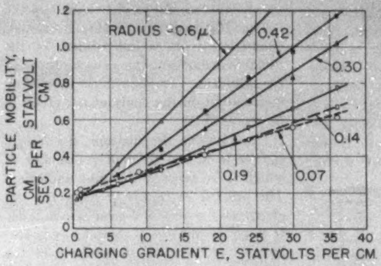


Fig. 7. Experimental results for Nt=10⁷, showing effects of charging gradient

aerosol of known particle size (equation 12). Or, if the charge is determined by mobility tests, this provides another method of measuring particle size.

The apparatus which was used for measuring the electric current carried by the aerosol stream is shown in Fig. 5. The charged particles in a measured flow of the aerosol are collected by the high-efficiency filter. The mass of liquid per unit volume of air is measured by the concentration meter which has been calibrated by precipitating the particles on aluminum foil and weighing. From these measurements, the charge per unit mass is determined.

Test Results

The particle sizes obtained from charge-mass measurements, equation 12, agree closely with those measured by the Owl. A third method of size determination is the use of the Stokes-Millikan relation, equation 14, along with the single-charge mobility. This was found to agree with the other two methods for particles larger than 0.28-micron radius. For smaller particles, a discrepancy appeared, reaching a maximum of about 12% at 0.15-micron radius. The Stokes-Millikan relation indicates a larger size than the other two methods. This raises the question of the applicability of the Millikan correction factor for dioctyl phthalate particles. The sizes given in this paper are those measured by the other two methods.

The experimental results presented here are for positive ion charging, that used in precipitators for air-conditioning applications.

The experimental points shown in Figs. 6 to 14, and 16, represent the average particle mobility or charge. Most of these points were obtained from mobility analyzer curves similar to that shown in Fig. 3. Some were obtained from charge-mass data, as indicated in the captions.

Fig. 6 shows that for each charging field, there is a particle size at which the mobility reaches a minimum. The minimum mobility increases rather rapidly with field strength. For practical charging conditions, the particles most difficult to precipitate are those between 0.1- and 0.2-micron radius.

In Fig. 7, the data are plotted to show directly the variation with field strength. For a given particle size, the intercept at $E=0$ indicates the mobility which could be obtained by thermal charging alone. The additional mobility produced by a charging field is approximately proportional to the strength of the field. It is evident that the field is an important factor even for particles as small as 0.07 micron.

Two mechanisms by which particles become charged by gas ions are field charging and thermal charging.[5–8] The electric field drives ions onto the particle. The second mechanism is that of diffusion

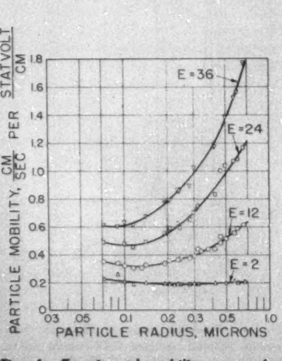


Fig. 6. Experimental mobility curves for Nt=10⁷. Note minimum mobilities

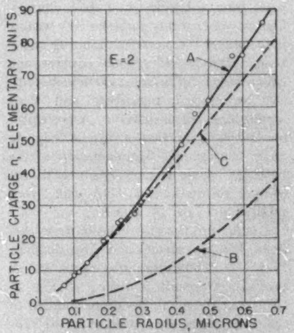


Fig. 8. Particle charge versus size for Nt=10⁷ and low gradient

A—Experimental results

B—Theoretical curve for field charging alone, by equation 19

C—Theoretical curve for thermal charging alone, by equation 23

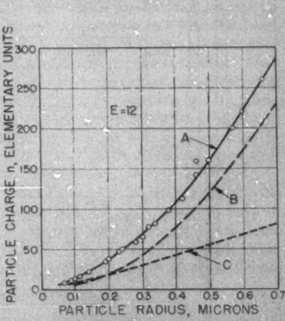


Fig. 9. Particle charge versus size for Nt=10⁷ and moderate gradient

For explanation of curves, see Fig. 8

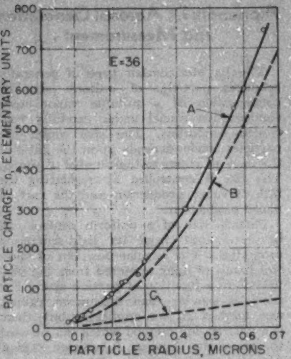


Fig. 10. Particle charge versus size for Nt=10⁷ and high gradient

For explanation of curves, see Fig. 8

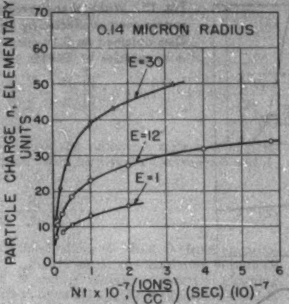


Fig. 12. Effect of Nt on charging of 0.14-micron particles. Data obtained by the mobility method for E=12, and by the charge-mass method for the other two curves

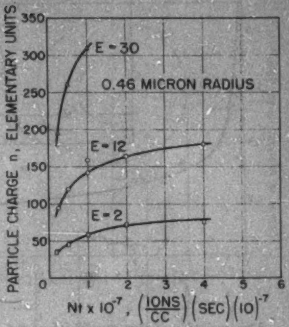


Fig. 14. Effect of Nt on charging of 0.46-micron particles. Data obtained by mobility method

of ions onto the particle because of their thermal energies. Theoretical equations, considering the two mechanisms separately, are given in Appendix V. Theoretical curves, from equations 19 and 23, are shown in some of the figures for comparison with the experimental curves. Addition of the field and thermal charges is not allowable because each equation was developed on the basis of only one mechanism acting. Rather, the charging rates, equations 18 and 22, should be added. The charge calculated from this is only slightly greater than the field or thermal charge, whichever is larger.

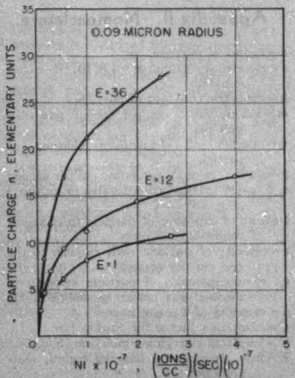


Fig. 11. Effect of Nt on charging of 0.09-micron particles. Data obtained by the charge-mass method

Figs. 8, 9, and 10 show particle charge as a function of size for low, medium, and high charging gradients. At low gradient, the experimental results agree reasonably well with theory. At the higher gradients, and for the smaller particles, the experimental charge is considerably greater than the theoretical charge, either field or thermal. For the larger particles, the discrepancy decreases.

The field charging equation 19 does not take into account the effect of the conductivity of the particle. Conductivity tends to permit the charge to distribute itself on the particle in such a way as to allow the particle to acquire a greater charge. The upper limit of this effect is equivalent to infinite dielectric constant in equation 16, which is a 40% increase in charge. Since the resistivity of dioctyl phthalate is only $3\,(10)^{10}$ ohm-centimeters (cm), the conductivity effect may account for an appreciable part of the discrepancy for the larger particles in Figs. 9 and 10.

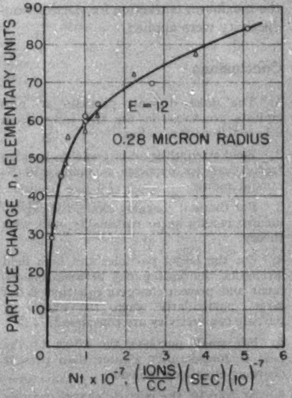


Fig. 13. Effect of Nt on charging of 0.28-micron particles

O = charge-mass data
Δ = mobility data

The variables N and t appear in the charging equations only as the product Nt. A test was made in which t was varied from 0.014 to 0.13 second, with Nt held constant at $0.25\,(10)^7$. There was no significant change in mobility for this variation in t and N. In another test, the mobility remained constant for variation of t from 0.07 to 0.13, at $Nt=2\,(10)^7$.

The charging time used in most of the tests was 0.07 second. For values of Nt greater than $2\,(10)^7$, it was necessary to increase t.

Table I. Effects of Nt

Particle Radius, Microns	Charging Gradient, E	Experimental Charge Ratios* A	Experimental Charge Ratios* B
0.07	2	0.81	1.23
	12	0.75	1.18
	30	0.72	1.15
0.14	1	0.80	1.22
	12	0.81	1.19
	30	0.80	1.19
0.28	12	0.83	1.18
	36	0.81	1.25
0.46	2	0.80	1.22
	12	0.83	1.16
	30	0.84	
0.66	12	0.82	1.17

* A is the ratio of the particle charge at $Nt=0.5\,(10)^7$ to that at $Nt=10^7$ and B is the ratio of the particle charge at $Nt=2\,(10)^7$ to that at $Nt=10^7$.

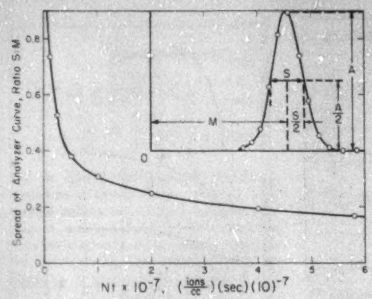


Fig. 15. Width of analyzer response curve as affected by Nt. Data obtained on 0.14-micron particles at E=12

The variation of charge with the product Nt is shown in Figs. 11 to 14. Note in Fig. 13 the close agreement of data obtained by the two independent methods, mobility measurements and charge-mass measurements.

These curves show a greater slope at $Nt=10^7$ than theory indicates. In Table I, A and B are, respectively, the ratios of the charge at $Nt=0.5\ (10)^7$ and $Nt=2\ (10)^7$ to that at $Nt=10^7$. Theoretically, for field charging $A=0.89$ and $B=1.06$, and for thermal charging A varies from 0.84 to 0.90 and B from 1.13 to 1.16 for this particle size range. The experimental curves do not flatten off as quickly as theory predicts, particularly field charging.

The effect of Nt upon the statistical variation of particle charge is indicated in Fig. 15. The actual spread in mobility is less than that of the analyzer response curve by a nearly constant amount which is attributable to conditions in the analyzer. A rough estimate of the amount to be subtracted from the ordinates of Fig. 15 to obtain spread in mobility, is 0.1. The statistical variation is small for large values of Nt. As Nt decreases below 10^7, the statistical spread increases rapidly.

Fig. 16 provides a comparison of experimental and theoretical mobilities at a moderate charging gradient. Curve D takes into account both charging mechanisms operating simultaneously by addition of the charging rates, equations 18 and 22. For the smaller particles, which includes the region of minimum mobility, the experimental mobilities are considerably greater than the theoretical, even if the maximum correction for particle conductivity were applied.

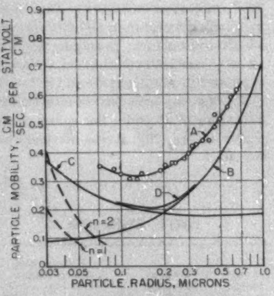


Fig. 16. Mobility curves for Nt=10⁷

A—Experimental data for E=12

B—Theoretical curve for field charging alone, E=12, by equation 19

C—Theoretical curve for thermal charging alone, by equation 23

D—Theoretical curve considering both field and thermal charging by addition of their rates given by equations 18 and 22

Broken lines show mobility of particles carrying one and two charges, according to equation 14

Conclusions

1. The most difficult particles to precipitate are those in the size range from 0.1- to 0.2-micron radius.

2. Field strength is an important charging factor even for particles as small as 0.07-micron radius.

3. For thermal charging alone, the experimental results agree reasonably well with theory.

4. For the larger particles and high field strengths, the discrepancy between experiment and present charging equations is not large, particularly when the effects of particle conductivity are considered.

5. For small particles the effect of the field is considerably greater than present charging equations indicate.

6. At $Nt=10^7$ the particle charge changes more rapidly with Nt than present charging equations predict.

7. The statistical variation of particle charge rises rapidly as Nt is decreased below 10^7.

Appendix I. Aerosol Generation and Measurement

The La Mer-Sinclair type of generator produces particles of uniform size by condensation of a suitable vapor upon condensation nuclei under carefully controlled conditions. The size which the particles attain depends upon the ratio of the mass of vapor to the number of nuclei. The size is controlled by regulating the rate of vapor production and the rate of nuclei production.

Particle size of a uniform aerosol can be determined from its light-scattering properties. The angular positions of spectral bands of light scattered from the particles, are a function of particle size. For particles smaller than 0.2-micron radius, this method is not applicable but there is a polarization method which can be used to extend the range to 0.1-micron radius. Also, by exposing the aerosol to the vapor of a suitable solvent under proper conditions, the particles can be "grown" by a known factor.[9] Then the size of the grown particles can be measured by the Owl and the original size calculated from the known growth factor. Particles smaller than 0.1-micron radius can be measured by this method.

The aerosol concentration meter is a forward-angle type of Tyndallometer which employs dark-field illumination in order to exclude direct light from the collecting lens system. There is provision for inserting an opal glass arbitrary standard into the light-scattering position when desired. The amount of light scattered depends upon particle size as well as upon particle concentration. This instrument was calibrated in terms of mass concentration, with respect to particle size, by electrostatically precipitating the particles from measured volumes of aerosols onto aluminum foil and weighing. The device for this is similar to that described by Barnes and Penney, except that an aluminum foil liner was used inside the cylinder.[10]

Appendix II. Nomenclature

a = radius of particle, cm
e = elementary charge = $4.8(10)^{-10}$ statcoulomb
k = dielectric constant of particle
m = ion mobility, cm per second per statvolt per cm
n = number of elementary charges on particle
q = air flow in analyzer, cubic cm per second
r, θ, z = cylindrical co-ordinates in analyzer (Fig. 18)
t = time of exposure of particle to charging conditions, seconds
v = velocity of particle with respect to the gas, cm per second
v_r, v_z = components of particle velocity in the analyzer, cm per second (Fig. 18)
μ = viscosity of the gas, poises
C = correction factor for Stokes law
D = length of charging region, cm (Fig. 1)
E = field strength in charging region, statvolts per cm
E_p = field strength in precipitating region, statvolts per cm
I = current to electrode A of Fig. 1, averaged over the whole cycle, statamperes

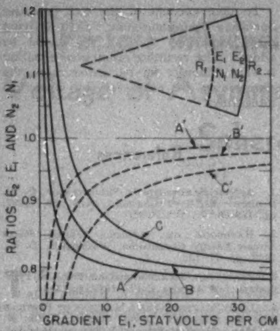


Fig. 17. Maximum variations of E and N in the particle charger. A, B, and C show the gradient variation, and A′, B′, and C′ show the ion density variation

A, A′—$N_1 = 0.7\,(10)^8$
B, B′—$N_1 = 1.4\,(10)^8$
C, C′—$N_1 = 2.8\,(10)^8$

K = Boltzmann constant = $1.38\,(10)^{-16}$
L = length of analyzing region, cm (Figs. 2 and 18)
M = particle mobility, cm per second per statvolt per cm
N = ion concentration in charging region, ions per cubic cm
Q = charge per gram of particles, statcoulomb per gram
R = radius of position in charger, cm (Fig. 17)
T = absolute temperature, degrees Kelvin
U = rms thermal velocity of ions, cm per second
V = analyzer voltage, statvolts (Fig. 2)

Appendix III. Aerosol Charger

The square wave applied between S and A actually is a clipped 60-cycle sine wave. By removing the upper portion of each half wave, flat tops extending 150 degrees with a voltage variation of 4%, are obtained. The 60-cycle frequency is satisfactory for the 0.72-cm spacing between S and A, and for the ranges of particle mobility and field strength under investigation.

Ions are present in C only during the half-cycle in which A is negative, so the effective charging time is 42% of the total time the particles spend in region C. If I is the current to electrode A, averaged over time, then

$$I = (0.42)(0.1)2\pi RDNemE \tag{1}$$

Substituting $m = 480$, the mobility of positive ions, N is calculated as follows:

$$N = 1.65(10)^7 \frac{I}{RDE} \tag{2}$$

The variation in E with radial position is easily obtained from Poisson's equation, expressed in cylindrical co-ordinates, for a long cylinder, in the following equation

$$\frac{1}{R}\frac{d(ER)}{dR} = 4\pi Ne \tag{3}$$

The solution of this equation is

$$\left[\frac{E_2}{E_1}\right]^2 = \left[\frac{R_1}{R_2}\right]^2\left[1 - 4\pi e\frac{R_1N_1}{E_1}\right] + 4\pi e\frac{R_1N_1}{E_1} \tag{4}$$

in which the subscripts refer to two positions, taken here as the electrodes S and A. Also

$$\frac{N_2}{N_1} = \frac{R_1}{R_2}\frac{E_1}{E_2} \tag{5}$$

for fixed values of I and D (from equation 2). Fig. 17 shows the maximum variations in E and N radially across the charging space, as calculated from equations 4 and 5.

The voltage gradient and ion concentration referred to in Figs. 6 to 16, are the average gradient in space C and the ion concentration at this average gradient assuming its position to be midway between S and A, as calculated from equation 2.

Appendix IV. Measurements of Mobility and Charge Per Unit Mass

As the charged particle is carried with velocity v_z axially by the air stream, it is forced by the electric field toward the inner cylinder at radial velocity v_r, as indicated in Fig. 18. Both of these components vary with radial position. The radial distribution of field strength and therefore of v_r is known, and that of v_z need not be known in order to calculate the particle mobility M.

$$v_z = \frac{dz}{dt} \tag{6}$$

$$v_r = -\frac{dr}{dt} = ME_p \tag{7}$$

and

$$E_p = \frac{V}{r \ln \frac{r_2}{r_1}} \tag{8}$$

where z and r are the axial and radial positions of the particle, E_p is the field strength at radius r, and V is the potential difference between cylinders.

$$dz = -v_z\frac{dr}{ME_p} = -v_z r\left[\ln\frac{r_2}{r_1}\right]\frac{dr}{MV}$$

from equations 6, 7, and 8, and

$$L = \frac{1}{MV}\ln\frac{r_2}{r_1}\int_{r_1}^{r_2} v_z r\,dr \tag{9}$$

But

$$dq = v_z 2\pi r\,dr$$

and

$$q = 2\pi\int_{r_1}^{r_2} v_z r\,dr \tag{10}$$

where q is the volume rate of air flow between cylinders. Substituting from equation 10 into equation 9,

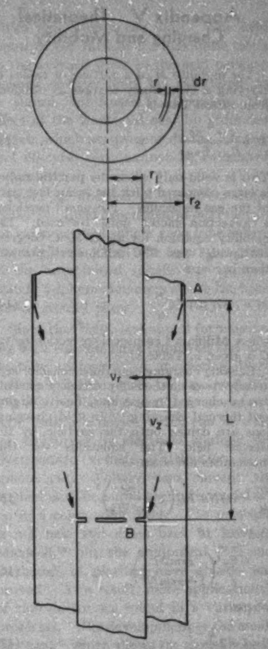


Fig. 18. Mobility analyzer particle entrance at A and sampling groove at B

$$L = \frac{1}{MV}\frac{q}{2\pi}\ln\frac{r_2}{r_1}$$

and

$$M = \frac{q}{2\pi LV}\ln\frac{r_2}{r_1} = 0.122\frac{q}{LV} \tag{11}$$

The spread in the mobility curve is not caused entirely by actual particle mobility distribution. Some broadening of the curve is caused by small eccentricity of the cylinders, finite widths of entrance and exit openings, blending of streams at the entrance, and some degree of turbulence and diffusion in the air stream. The spread of the curve was found to be affected very little by L and q in the ranges used. A measure of the spread is provided by dividing the width of the curve at the half peak ordinate, by the average mobility, as indicated in Figs. 3 and 15. The smallest spread obtained had a ratio of 0.14, so the spread caused by conditions in the analyzer must be less than this.

In the charge-mass measurements on dioctyl phthalate particles of uniform size, the relation is simply

$$Q = \frac{3ne}{4\pi(0.984)a^3} = 1.17(10)^{-10}\frac{n}{a^3} \tag{12}$$

Appendix V. Theoretical Charging and Mobility

The mobility M of a particle of radius a, carrying n elementary charges as obtained from Stokes Law is

$$M=\frac{v}{E_p}=\frac{ne}{6\pi\mu a} \tag{13}$$

This is valid only when the particle radius is large compared with the mean free path of the gas molecules. For small particles, a correction factor must be used. The mobility equation for air at 30 C (degrees centigrade) and 730 millimeters pressure then becomes

$$M=1.37(10)^{-7}C\frac{n}{a} \tag{14}$$

where Millikan's empirical correction factor C is given in Table II.[11]

Existing charging equations consider separately two methods by which a particle can be charged by gas ions, field charging and thermal charging.[5-8] In field charging, ions are driven onto the particle by the electric field. The equations for this mechanism alone are

$$\frac{dn}{dt}=\pi mNa^2EG\left(1-\frac{ne}{a^2EG}\right)^2 \tag{15}$$

and

$$n=\frac{GEa^2}{e}\left(\frac{\pi emNt}{1+\pi emNt}\right) \tag{16}$$

where

$$G=1+2\,\frac{k-1}{k+2} \tag{17}$$

k is the dielectric constant of the particle and m is the mobility of the ions. For $k=5.1$, the dielectric constant of dioctyl phthalate, and for positive ions in air at 30 C and 730 millimeters pressure,

$$\frac{dn}{dt}=3{,}250\ NEa^2\left[1-\frac{2.22(10)^{-10}n}{Ea^2}\right]^2 \tag{18}$$

and

$$n=45(10)^8Ea^2\left[\frac{7.24(10)^{-7}Nt}{1+7.24(10)^{-7}Nt}\right] \tag{19}$$

In thermal charging, the ions reach the particles by virtue of their thermal energies. Considering thermal charging alone, the equations given by White are

$$\frac{dn}{dt}=\pi UNa^2\epsilon^{-\frac{ne^2}{aKt}} \tag{20}$$

$$n=\frac{aKT}{e^2}\ \ln\left(1+\frac{\pi Ue^2aNt}{KT}\right) \tag{21}$$

where U is the rms thermal velocity of the ions. For air at 30 C, assuming the ions to consist of single molecules in thermal equilibrium with uncharged molecules,

$$\frac{dn}{dt}=15.7(10)^4Na^2\epsilon^{-5.56(10)^{-6}\frac{n}{a}} \tag{22}$$

and

$$n=1.8(10)^5a\ \ln\,(1+0.9aNt) \tag{23}$$

To take account of the two processes operating simultaneously, the charging rates, equations 19 and 22, should be added. The resulting differential equation is very difficult; for particular cases, the charge can be calculated by numerical integration.

Table II. Correction Factor for Stokes Law

Particle Radius a	 Cm	Correction Factor, C
0.07	$(10)^{-4}$	2.40
0.10		1.94
0.15		1.62
0.20		1.45
0.30		1.28
0.40		1.22
0.60		1.15
0.80		1.11
1.00		1.09

References

1. A NEW ELECTROSTATIC PRECIPITATOR, **G. W. Penney.** *AIEE Transactions,* vol. 56, 1937, pp. 159–63.

2. HANDBOOK ON AEROSOLS. United States Atomic Energy Commission, Washington, D. C., 1950, pp. 77–102.

3. *Proceedings,* First National Air Pollution Symposium, Los Angeles, Calif., 1949, pp. 1–13.

4. LIGHT SCATTERING AS A MEASURE OF PARTICLE SIZE IN AEROSOLS, **David Sinclair, Victor K. La Mer.** *Chemical Reviews,* Baltimore, Md., vol. 44, 1949, pp. 245–67.

5. PARTICLE CHARGING IN ELECTROSTATIC PRECIPITATION, **H. J. White.** *AIEE Transactions,* vol. 70, pt. II, 1951, pp. 1186–91.

6. CHARGING OF SPHERICAL PARTICLES IN AN IONIZING FIELD, **M. M. Pauthenier, M. Moreau-Hanot.** *Journal de Physique et Le Radium,* Paris, France, vol. 3, Ser. 7, 1932, pp. 590–613.

7. RESEARCH ON THE PHYSICAL BASIS OF ELECTRICAL GAS PURIFICATION, **R. Ladenburg.** *Annalen der Physik,* Leipzig, Germany, vol. 4, 1930, pp. 863–97.

8. THE MECHANISM OF CHARGING OF MIST PARTICLES, **P. Arendt, H. Kallman.** *Zeitschrift fur Physik,* Berlin, Germany, vol. 35, 1925, pp. 421–41.

9. A DIRECT TEST OF KELVIN'S EQUATION CONNECTING VAPOR PRESSURE AND RADIUS OF CURVATURE, **Victor K. La Mer, Ruth Gruen.** *Transactions,* Faraday Society, London, England, vol. 48, 1952, pp. 410–15.

10. AN ELECTROSTATIC DUST WEIGHT SAMPLER, **E. C. Barnes, G. W. Penney.** *Journal of Industrial Hygiene and Toxicology,* Chicago, Ill., vol. 20, no. 3, March 1938, pp. 259–65.

11. GENERAL LAW OF FALL OF A SMALL SPHERICAL BODY THROUGH A GAS, **R. A. Millikan.** *Physical Review,* New York, N. Y., vol. 22, Ser. 2, 1923, pp. 1–23.

◆

Discussion of 57-89 and 57-90

G. W. Hewitt and G. W. Penney: Measurements of the charges on particles whose diameters are smaller than the wavelength of visible light, must be made indirectly. There are various possible sources of error. For this reason, there is some degree of uncertainty in the results, and it is particularly important to compare the results of independent investigations.

The experimental results presented in the two papers are rather difficult to compare because there are several parameters; different ranges are covered; and the data are plotted in different ways. The Penney-Lynch (57-89) investigation was directed primarily toward the industrial precipitator which involves large values of Nt, while the Hewitt work was directed more toward the low-voltage 2-stage precipitator with its lower values of Nt. However, there are regions of overlap in which comparisons can be made. It is difficult for the casual reader to compare the results, so it seems desirable to point out the degree of agreement.

Since, in the two papers, there are no points for which all parameters are exactly the same, points were chosen for which the parameters are nearly the same and corrections were made for the parameter differences. For example, in Fig. 11 of the Penney-Lynch paper, the positive particle charge at $Nt=6\ (10)^7$ and 11 kv (for which Fig. 5 shows the field strength to be 2.7 kv per cm), for particles of 0.15-micron radius, is 33. In comparison, Fig. 12 of the Hewitt paper (57-90) shows 34 charges for a 0.14-micron particle at $Nt=6\ (10)^7$ and $E=12$ statvolts per cm or 3.6 kv per cm. Correcting for the size difference by using Fig. 9 of paper 57-90, and for difference in field strength by using Fig. 7, the indicated charge becomes 33, the same as the value given by paper 57-89. Using this method of comparison, values from various curves of paper 57-89 compare with the Hewitt data as follows.

Particle charge values shown in the lower half of Fig. 8 and in the whole range of Fig. 9, are about 10% lower than values obtained from paper 57-90. In Fig. 10, comparisons can be made in the lower portion of the positive curve. The values shown are lower than the Hewitt data by less than 10%.

It is only in Fig. 7 that the two sets of data differ by more than 10%. In this case, the Penney-Lynch values of charge are higher than the Hewitt values; the difference varies from 15% at 0.16-micron to 27% at 0.28-micron radius. Except for these discrepancies in Fig. 7, the agreement between the two sets of data is good, considering the difficulties in this type of experiment.

Although there are small differences in the results, both papers indicate that the charge acquired by small particles in the electric fields commonly used in electrostatic precipitation, are considerably higher than those predicted by existing charging equations, especially at large values of Nt.

Factors Influencing the Sparkover Voltage of Asymmetrically Connected Sphere Gaps

H. E. FIEGEL
ASSOCIATE MEMBER AIEE

W. A. KEEN, JR.
ASSOCIATE MEMBER AIEE

THE sphere gap which is widely used as a method of measuring high voltage has its acceptance based on extensive calibration of the breakdown of air as a dielectric between spheres of specified sizes and spacings. The breakdown voltages between spheres of particular sizes are given for spacings up to a sphere diameter in the AIEE Standard no. 4[1] and in the British Standard no. *358*.[2]

The use of the grounded sphere gap for high-voltage measuring purposes was first proposed to the AIEE in 1913 by Chubb and Fortescue,[3] and Peek.[4] Peek derived an empirical equation based on his own measurements for calculating the sparkover of sphere gaps of different diameters and spacing. In 1914, the AIEE adopted calibration curves for sphere gaps having spheres of 6.25, 12.5, 25, and 50 cm (centimeters) in diameter. With the advent of impulse testing and the much higher voltages available, spheres of 100, 150, and 200 cm in diameter came into use. It soon became apparent[5] that gaps with different sizes of spheres could not be checked accurately against each other. Also, it was noted that there was a marked polarity effect[6,7] at certain spacings and that there was a wave-shape effect when the gaps were subjected to very steep impulse waves.[8] At about this same time in a CIGRE[9] session, data were presented on the influence of nearby vertical ground planes on the sparkover voltage of grounded sphere gaps. In 1934 considerable data on the calibration of sphere gaps were published by Meador,[10] and Bellaschi and McAuley.[11] The present accepted sparkover voltages, as published by AIEE, and the empirical formulas developed by Ver Planck[12] were both based on these data. Throughout this period, when the sphere gap was being accepted and calibrated, and in the more recent publications on its use,[13–20] very little has been presented as to the relative errors introduced by the many field-distorting influences.

Errors of voltage measurements as stated in the standards will be ±3% when spheres are used in accordance with certain recommendations. These recommendations[1] pertain to corrections for relative air density, sphere and sphere frame construction, gap irradiation, and precautions concerning external electric fields and ground effects. One of these precautions states that, "the sphere gap should be located in a space comparatively free from external electric fields" and then gives a distance approaching twice the sphere diameter as the minimum. The precaution which relates to the effect of ground says that "variations in sparkover voltage due to the effect of ground, adjacent objects, walls and other field-distorting influences may require consideration especially for gap spacings above one half the sphere diameter." The recommended distances to ground are between 5 and 10 diameters for 50-cm and smaller sphere gaps and proportionally less for the larger sizes, and as low as 2.5 diameters for the 200-cm spheres. These precautions in themselves are not enough to hold the accuracy within the limits of error prescribed. It is the purpose of this paper to present realistic field factors based on the proximity of the horizontal ground plane under the spheres and to indicate the sparkover variations caused by other field-distorting influences.

Field Factor

To understand how the individual influences contribute to the sparkover voltages of sphere gaps it is necessary to know how they affect the distribution of voltage in the space between the spheres and particularly how they affect the regions which are stressed the most. The field factor or the ratio of the maximum stress to the average stress in the field between isolated spheres can be calculated as is shown by Russel.[21] However, the effect of the leads and near-by grounded objects which are present during impulse and 60-cycle tests cannot be calculated accurately.

From an analysis of sparkover data, Peek[22] obtained experimental field factors for asymmetrical sphere gaps (one sphere grounded) in which the sparking point of the ungrounded sphere was approximately 5 sphere diameters above the horizontal ground plane.

Since that time, techniques for measuring high voltages have improved and conduction analogues for measuring these ratios of gradients have become available. The electrolytic field analyzer[23] or electrolytic tank is a device which can be used for measuring by electrical analogy the electrostatic voltage distribution in a sphere gap. Voltage and current are applied and measured around a model set up in a conducting fluid. The principle is not new and dates back at least to Maxwell,[24] but the equipment [25,26] and methods[27] of plotting have greatly improved. As a result, more applications[28] of this device are found in a variety of problems. The development of the auxiliary tank[25] which allows the model boundaries, in effect, to extend to infinity makes possible the representation of electric field influences which are relatively large distances away from the stressed region under consideration.

The sloping-bottom electrolytic tank in conjunction with its auxiliary tank was used because of the rotational symmetry of the sphere gap. Gradient measurements along the center line of the gap (shore line of the model) were desired as this coincides with the flux line which originates at the point of maximum gradient. However, since the electrolyte depth is zero along this line, such measurements were made along flux lines approximately parallel and as close as practical to the center line. Comparison of the measured surface gradient with the exact calculated maximum gradient, for a sphere gap with both spheres insulated, indicated a high accuracy of this method of measurement.

Using this same technique, the effect of certain physical factors such as shank size *d*, sphere diameter *D*, gap spacing *S*, and distance from sparking point to horizontal ground plane *P* were determined for vertically mounted asymmetrical sphere gaps. The field factors *f* for a gap spacing of 1 sphere diameter with var-

---

Paper 57-20, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 23, 1956; made available for printing November 20, 1956.

H. E. Fiegel and W. A. Keen, Jr., are with the General Electric Company, Pittsfield, Mass.

The authors wish to acknowledge the help of L. J. Wiesstort and all other personnel of the General Electric High-Voltage Laboratory who assisted in gathering these data. In particular, acknowledgment is made of the guidance and direction of J. H. Hagenguth and A. F. Rohlfs, and the valuable suggestions of H. K. Farr.

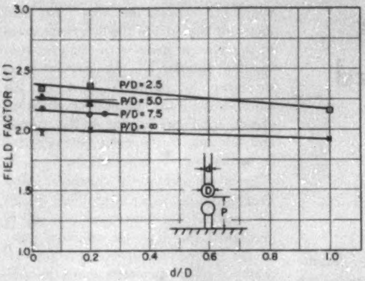


Fig. 1 (left). Effect of shank diameter on field factor

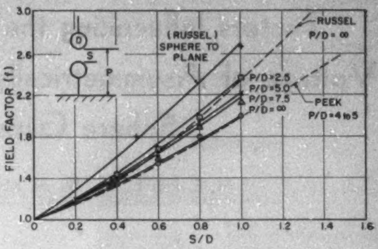


Fig. 2 (right). Field factors

ious shank sizes are plotted in Fig. 1, as f versus d/D for several positions of the sphere gap above the horizontal ground plane. Examination of these field-factor curves reveals that within practical limits, from one fifth to one twentieth of the sphere diameter, the effect of shank size is negligible.

The field factor versus the ratios of gap spacing to the sphere diameter S/D are plotted in Fig. 2, with P/D as the parameter for asymmetrically connected sphere gaps or sphere gaps with one sphere grounded. The field-factor curve for a sphere-to-plane geometry would have to fall on or above a curve for $P/D=1.0$ in the previously mentioned graph. An exact field calculation, one in which no approximations are necessary, was made by Russel[13] for this configuration, and his curve as well as measured values have been added to Fig. 2.

It was also noted that, when the reciprocal of the field factor versus $(D/P)^{1/2}$ was plotted, a family of straight lines were obtained with S/D as the parameter. From this family of lines, shown in Fig. 3, the equation

$$1/f = A - B(D/P)^{1/2}$$

was derived. The values of the parameters A and B for various gap spacings are plotted in Fig. 4.

High-Voltage Test Setup

To determine the field influence of one variable at a time on sparkover voltages under actual test conditions, high-voltage tests were performed in a relatively large test area, the floor and elevation plan of which are shown in Fig. 5. The ceiling (overhead influencing ground structures) over the area was 75 feet above the floor, except when the gaps were elevated in air. For these tests, a bridge crane was used to support the gaps and the lowest plane of the crane's bridge was 65 feet above the floor and was 20 feet wide. The crane hook, in turn, was located 15 feet minimum above the 50-cm spheres and 26 feet above the 100-cm spheres. The circuit leads to the impulsed (top) sphere came in horizontally to a vertical extension of the sphere shank, 6 feet above the sparking surface of the upper sphere. The impulse generator was 44 feet high and the upper 26 feet rose to line potential for most of the tests on the 50-cm sphere gaps. The generator was 17 banks high, but for these tests only 5 or 10 banks were used. For these generator connections, the idle banks were connected electrically to the top bank of the active generator and, thus, when the generator was discharged they assume the output voltage of the generator.

All tests reported were made using nominal 1.5x40-microsecond waves of negative polarity. The minimum and maximum front times were 1.5 and 1.9 microseconds respectively, with minimum and maximum tail times of 40 and 50 microseconds. Voltage measurements were made by use of a resistance voltage divider of 5,000 or 10,000 ohms and a cathode-ray oscillograph. The critical sparkover voltages were obtained by applying 10 impulses per level, starting at a withstand level and increasing the voltage in about 1.0% steps until a level was reached which resulted in all sparkovers. The critical level was determined by plotting voltage against frequency of sparkover and reading the voltage which would have resulted in sparkover 50% of the time. In 81% of the cases, the number of levels required between the all-withstand and the all-sparkover level was three or less, while 53% required only two levels. The variation in voltage between impulses in one level was about 1%, and the over all spread in voltages to obtain the critical sparkover was about 2% for more than 50% of the cases.

High-Voltage Tests

HORIZONTAL GROUND PLANE INFLUENCE

To determine how well the analyzer field factor correlated with actual test results, high-voltage tests were made on a set of vertically mounted 50-cm sphere gaps located at various distances above a horizontal ground plane, with sphere spacings of 1.0, 0.6, and 0.4 diameter. The 50-cm spheres were chosen because, from the literature, they appeared typical of the greatest number of laboratories. They also (again from the literature) appeared to be of a size which produced

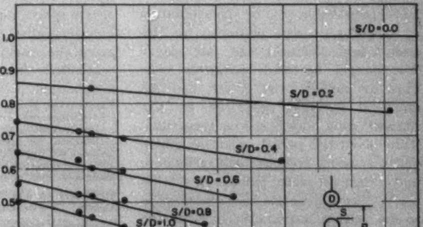


Fig. 3. Linear relation between field factor and sphere-gap geometry

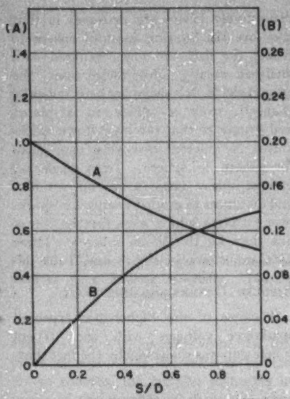


Fig. 4. Variations of parameters in equation for the field factor, $1/f = A - B(D/P)^{1/2}$

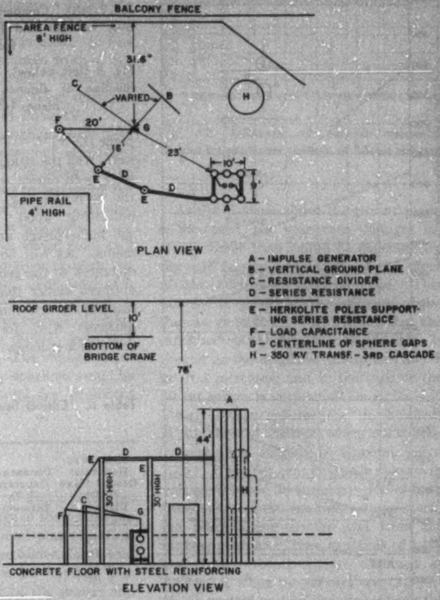

BALCONY FENCE
AREA FENCE
8' HIGH
31.6"
C
VARIED
B
H
F
20'
G
16'
23'
10'
9'
E
D
E
D
A
PIPE RAIL
4' HIGH
PLAN VIEW
A - IMPULSE GENERATOR
B - VERTICAL GROUND PLANE
C - RESISTANCE DIVIDER
D - SERIES RESISTANCE
E - HERKOLITE POLES SUPPORTING SERIES RESISTANCE
F - LOAD CAPACITANCE
G - CENTERLINE OF SPHERE GAPS
H - 350 KV TRANSF - 3RD CASCADE
ROOF GIRDER LEVEL
10'
BOTTOM OF BRIDGE CRANE
78'
A
44'
E
D
D
E
30 HIGH
30 HIGH
C
F
G
CONCRETE FLOOR WITH STEEL REINFORCING
ELEVATION VIEW

Fig. 5. Plan and elevation view of high-voltage test area

fairly good correlation between laboratories, and they were sufficiently small to permit large changes in distances to influencing structures. The results of these tests are plotted in Fig. 6 and listed in Table I, with the various distances given in diameters.

From the curves note that the voltage increases with increasing values of P/D. The field factors obtained by the analyzer had indicated that an increase would be obtained; i.e., as the field factor for a particular spacing is decreased the sparkover voltage will increase. If this voltage increase is directly proportional to the reciprocal of the field factor, the critical sparkover gradient for a given sphere diameter and gap spacing must be constant, and the sparkover curves for various values of $(D/P)^{1/2}$ should be straight lines. These lines should also have the same slopes as were obtained in the plot of $1/f$ versus $(D/P)^{1/2}$ in Fig. 3.

In Fig. 7 the various critical sparkover voltages are plotted versus the square root of the ratio of sphere diameter to the distance to the horizontal ground

Table I. Critical Sparkover Voltages as Function of Distance to Horizontal Ground Plane for Set of 50-Cm Spheres

Distance to Horizontal Ground Plane (P/D), Diameters	Distance to Generator (L/D), Diameters	Distance to Resistance Divider, Diameters	Distance to Vertical Ground Plane, Diameters	AIEE Crest Voltage, Kv	Test* Crest Voltage, Kv	Per-Cent Difference
			Gap Spacing S/D = 1.0			
14	14	≥12.2	≥12.2	707	794	+12.3
9.25	14	≥12.2	≥12.2	707	749	+ 5.95
6.97	14	≥12.2	≥12.2	707	726	+ 2.7
4.68	14	≥12.2	≥12.2	707	706	− 0.14
4.2	14	≥12.2	≥12.2	707	699	− 1.13
2.16	14	≥12.2	≥12.2	707	657	− 7.07
1.0	14	≥12.2	≥12.2	707	581	−17.8
			Gap Spacing S/D = 0.6			
14	14	≥12.2	≥12.2	573	601	+ 4.9
9.25	14	≥12.2	≥12.2	573	587	+ 2.44
7.02	14	≥12.2	≥12.2	573	581	+ 1.4
4.68	14	≥12.2	≥12.2	573	571	− 0.3
4.2	14	≥12.2	≥12.2	573	569	− 0.7
1.76	14	≥12.2	≥12.2	573	541	− 5.6
0.6	14	≥12.2	≥12.2	573	467	−18.5
			Gap Spacing S/D = 0.4			
14	14	≥12.2	≥12.2	451	465	+ 3.1
9.2	14	≥12.2	≥12.2	451	460	+ 2.0
7.02	14	≥12.2	≥12.2	451	458	+ 1.55
4.2	14	≥12.2	≥12.2	451	450	− 0.22
1.56	14	≥12.2	≥12.2	451	434	− 3.75
0.40	14	≥12.2	≥12.2	451	373	−17.2

* Corrected for relative air density.

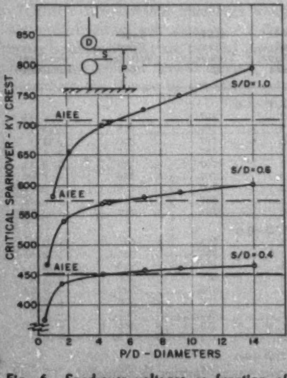


Fig. 6. Sparkover voltages as function of distance to horizontal ground plane, 50-cm spheres

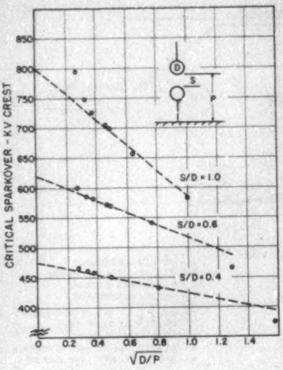


Fig. 7 (left). Comparison of variation of sparkover voltages and field factors for 50-cm spheres

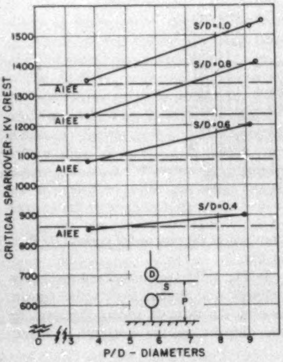


Fig. 8 (left, below). Sparkover voltages as function of distance to horizontal ground plane; 100-cm spheres

plane $(D/P)^{1/2}$. The data plotted follow fairly well the straight dashed lines which have been drawn with the same slope as the corresponding S/D lines of Fig. 3, indicating good agreement between the high-voltage and analyzer measurements. As will be noted at the larger distances to the ground plane, sparkover voltages higher than the predicted ones resulted, which indicated additional influences. Also, for P/D equal to 0.6 and 0.4 for spacings of 0.6 and 0.4 diameter respectively (these are extremes in that they are the limiting cases of sphere to plane) the data are low compared with analyzer results. For these cases, the ground electrode's shape has been radically changed, from a sphere to a plane, and it may be that the breakdown mechanism is different for this case than for sphere to sphere. Tests made on 100-cm sphere gaps of various spacings and distances to ground resulted in sparkover voltages which were further out of line than for the 50-cm spheres. These data are shown in Fig. 8 and Table II.

Impulse Generator Influence

A cause of the higher-than-expected sparkover voltages with the 100-cm sphere gap was suggested by the fact that,

Table II. Critical Sparkover Voltages as Function of Distance to Horizontal Ground Plane for 100-Cm Spheres

Distance to Horizontal Ground Plane (P/D), Diameters	Distance to Generator (L/D), Diameters	Distance to Resistance Divider, Diameters	Distance to Vertical Ground Plane, Diameters	AIEE Crest Voltage, Kv	Test* Crest Voltage, Kv	Per-Cent Difference
Gap Spacing S/D = 1.0						
3.71	7	6.1	8.54	1,338	1,351	+ 1.0
9.15	7	6.1	8.54	1,338	1,530	+14.2
9.53	9	4.46	7.5	1,338	1,550	+15.9
Gap Spacing S/D = 0.8						
3.71	7	6.1	8.54	1,234	1,234	0
9.33	9	4.46	7.5	1,234	1,410	+14.25
Gap Spacing S/D = 0.6						
3.71	7	6.1	8.54	1,084	1,081	− 0.28
9.13	9	4.46	7.5	1,084	1,201	+10.8
Gap Spacing S/D = 0.4						
3.71	7	6.1	8.54	862	853	− 1.04
8.93	9	4.46	7.5	862	902	+ 4.63

* Corrected for relative air density.

Table III. Critical Sparkover Voltages as Function of Distance to Impulse Generator

Distance to Generator (L/D), Diameters	Distance to Horizontal Ground Plane (P/D), Diameters	Distance to Resistance Divider, Diameters	Distance to Vertical Ground Plane, Diameters	AIEE Crest Voltage, Kv	Test* Crest Voltage, Kv	Per-Cent Difference
50-Cm Spheres, S/D = 1.0						
14	14	>12.2	>12.2	707	795	+12.45
9.15	14	>12.2	>12.2	707	818	+15.7
4.57	14	>12.2	>12.2	707	856	+21.1
14	7.02	>12.2	>12.2	707	717	+ 1.7
9.15	7.02	>12.2	>12.2	707	729	+ 3.11
4.57	7.02	>12.2	>12.2	707	750	+ 6.09
14	4.2	>12.2	>12.2	707	708	+ 0.14
9.15	4.2	>12.2	>12.2	707	710	+ 0.42
4.57	4.2	>12.2	>12.2	707	712	+ 0.707
50-Cm Spheres, S/D = 0.6						
14	14	>12.2	>12.2	573	601	+ 4.9
9.15	14	>12.2	>12.2	573	622	+ 8.55
4.57	14	>12.2	>12.2	573	642	+12.0
14	7.02	>12.2	>12.2	573	581	+ 1.4
9.15	7.02	>12.2	>12.2	573	592	+ 3.31
4.57	7.02	>12.2	>12.2	573	597	+ 4.2
14	4.2	>12.2	>12.2	573	569	− 0.7
9.15	4.2	>12.2	>12.2	573	574	+ 0.17
4.57	4.2	>12.2	>12.2	573	577	+ 0.7
100-Cm Spheres, S/D = 0.6						
7	3.71	> 6.1	> 6.1	1,084	1,080	− 0.37
4.57	3.71	> 6.1	> 6.1	1,084	1,102	+ 1.66
2.28	3.71	> 6.1	> 6.1	1,084	1,115	+ 2.86

* Corrected for relative air density.

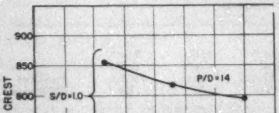


Fig. 9. Sparkover voltages as function of distance from impulse generator; 50-cm spheres

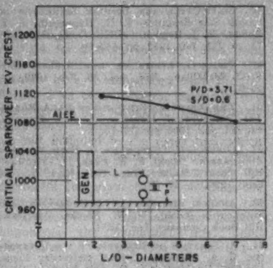


Fig. 10. Sparkover voltages as function of distance from impulse generator, 100-cm spheres

even though the distance from the impulse generator was the same as for the 50-cm spheres, the relative distances in terms of the number of diameters were cut in half. To determine the effect produced by the impulse generator proximity, the 50-cm sphere gaps were tested using spacings of 1.0 and 0.6 diameter at distances to the horizontal ground of 4.2, 7.02, and 14 diameters, while varying the distance to the generator *L*. The results of these tests are plotted in Fig. 9 and listed in Table III. The results of tests on a 100-cm sphere gap at a spacing of 0.6 diameter at a distance above ground of 3.71 diameters and at various distances from the generator are shown in Fig. 10 and Table III. From the plots and the tables, note that the influence of the generator becomes rather pronounced especially at large gap spacings and large distances to the horizontal ground plane.

Vertical Ground Plane and Resistance Divider Influences

Other rather obvious influences which are present in all test areas where floor space is rather limited are vertical ground planes and voltage divider effects. To determine the change in sparkover voltage introduced by varying the distance to large vertical ground planes, a 17.5-foot-high and 12-foot-wide vertical ground screen of chicken wire was used. It was located at various distances from a 50-cm sphere gap with 1.0- and 0.6-diameter spacings. The results of these tests are presented in Fig. 11 and Table IV.

The divider used throughout these tests were made from wire-wound noninductive resistance cards. The height of the divider used for the tests to determine the influence of the voltage divider proximity was about 17 feet. For this set of tests, the divider position was varied from 12.2 diameters in to 3.05 diameters from the center line of the gap. The tests at a 1.0-diameter gap spacing produced a change in the sparkover voltage of about 1% when moving the divider from 12.2 in to 3.05 diameters. The change introduced with smaller gap spacings would be even less. These results are presented in Fig. 12.

Discussion

The data show that the major influences on the sparkover voltage of the sphere gap are the horizontal ground plane, vertical ground plane, and large high-voltage conducting parts, such as an impulse generator. The influence of the shanks and the voltage divider are minor. Only the effect of a large horizontal ground plane was investigated with two independent methods; the 2-dimensional electrolytic field analyzer, and impulse sparkover tests in a large test area on a 50-cm sphere. As Fig. 7 shows, there is good correlation between the two sets of data for $(D/P)^{1/3}$ in the range of 0.4 to 1 ($P/D=1$ to 7), indicating that other influences were small. Under such conditions the sparkover voltage will vary inversely as the field factor for a given spacing and diameter, as given by the equation

$$1/f=A-B(D/P)^{1/3}$$

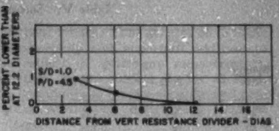


Fig. 12. Influence of vertical resistance divider on sparkover voltage of 50-cm spheres

As the distance above the ground plane is increased, the sparkover voltages increase faster than would be expected by the field factors established for the effect of the horizontal ground plane. This increase is caused by the influence of the impulse generator. Fig. 9 shows that the influence of the high-voltage field is greater as the height of the sphere above ground increases, and in the case of the 50-cm sphere is very small for $P/D=4.2$. For the 100-cm spheres (Fig. 10) the effect is apparently greater when comparing values on a distance-to-diameter basis, L/D. If the per-cent change in sparkover voltage is based on equal position of both the 50-cm and 100-cm high-voltage sparking point above ground and from the impulse generator, then increases of 3.1% and 3.5% respectively are obtained at $S/D=0.6$, when the spheres are 12.2 feet above ground and 6.5 feet away from the generator. This of course means that the high-voltage field effects will be greater for the larger spheres because of their physical height, since even at small values of P/D the sparking point will be closer to the top of the generator. When assessing the effect of extreme high-voltage fields on sphere sparkover, this relation must be taken into account.

The influence of a vertical ground plane of limited dimensions (12 feet wide and 17.5 feet high) is in the same direction as the horizontal ground plane, resulting in a reduction in sphere sparkover voltage as the ground is moved toward the spheres

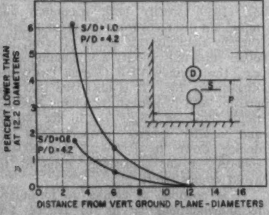


Fig. 11. Influence of vertical ground plane on sparkover voltages of 50-cm spheres

Table IV. Critical Sparkover Voltages as Function of Distance from Vertical Ground Plane for Set of 50-Cm Spheres

Distance to Vertical Ground Plane, Diameters	Distance to Horizontal Ground Plane (P/D), Diameters	Distance to Generator (L/D), Diameters	Distance to Resistance Divider, Diameters	AIEE Crest Voltage, Kv	Test* Crest Voltage, Kv	Per-Cent Difference
			Gap Spacing S/D = 1.0			
12.2	4.2	14	12.2	707	699	−1.13
6.1	4.2	14	12.2	707	689	−2.55
3.05	4.2	14	12.2	707	656	−7.2
			Gap Spacing S/D = 0.6			
12.2	4.2	14	>12.2	573	585	+2.09
6.1	4.2	14	>12.2	573	582	+1.57
3.05	4.2	14	>12.2	573	575	+0.35

* Corrected for relative air density.

Table V. Summary of Relative Influences

Sphere Size	P/D	L/D	Per-Cent Deviations in Sparkover Voltage From Those Given in Standards Table			
			S/D = 1.0	S/D = 0.8	S/D = 0.6	S/D = 0.4
50	5	14*	0		0	0
50	10	14*	+ 6.9		+ 2.6	+2.2
50	5	4.5*	+ 2.1		+ 1.4	
50	10	4.5*	+12.4		+ 7.8	
50	5	5†	− 2.3		− 0.8	
100	4.2	7*	+ 2.0	+ 1.0	+ 0.7	−0.8
100	10	9*	+16.6	+15.9	+12.5	+5.5
100	3.7	2.25*			+ 2.9	

AIEE Standard no. 4 precautionary remarks: distance to external electric high-voltage fields, $L > 2$ diameters; distance to ground, 5 to 10 diameters for 50-cm and smaller spheres down to 2.5 diameters for 200-cm spheres.

* L = distance to the generator. † L = distance to vertical ground plane.

(Fig. 11). It is very pronounced at large spacings, but its effect becomes small at $S/D = 0.6$ or less.

Many of these influences have been discussed in the past, and certain precautions have been written in the various standards to minimize deviations in sparkover voltages from those given in the standards' tables during normal testing procedures. Table V has been prepared to indicate the possible variations from standard tabular values if the precautions given on distances between spheres and field-influencing objects, given in AIEE Standard no. 4, are used. The table indicates that these limits are too generous and can result in considerably greater errors than 3%. This is particularly true for spacings in excess of $S/D = 0.6$.

In general, it is indicated that the sparking point of the high-voltage sphere should be in the range from 3 to 5 diameters above ground for 50-cm and larger spheres, and distances to other grounds should be of the same order or greater. The distance to high-voltage carrying parts should also be in excess of these distances to minimize the deviation in sparkover voltages from those given in the tables.

Conclusions

1. Realistic field factors that vary linearly as $(D/P)^{1/4}$ were determined which include the influences of the shank diameter and location of the horizontal ground plane.

2. It was shown that, while adhering to precautions stated in the standards pertaining to the use of asymmetrical sphere gaps, variations in sparkover voltages, as indicated in the following, can be obtained as a result of changes in the distance to:
   a. Horizontal ground plane...>3%.
   b. Impulse generator...>3%.
   c. Vertical ground plane...≤3%.
   d. Vertical resistance divider...<1%.

3. The influence of shank size is negligible when the shank diameters are kept equal to or less than one-fifth sphere diameter.

4. A possible revision in the precautions to be taken in the use of sphere gaps as a high-voltage measurement standard should be investigated.

References

1. Measurement of Voltage in Dielectric Tests. *AIEE Standard no. 4* and *ASA Standard no. C68.1*, American Standards Association, New York, N. Y., Sept. 1953.

2. Rules for the Measurement of Voltage With Sphere-Gaps. *Standard No. 358*, British Standards Institution, London, England, Dec. 1939.

3. Calibration of the Sphere Gap Voltmeter, L. W. Chubb, C. Fortescue. *AIEE Transactions*, vol. 32, 1913, pp. 739–48.

4. *Discussion* by F. W. Peek, Jr. of reference 3. *Ibid.*, p. 812.

5. Better Standards Needed for Surge Testing, F. D. Fielder. *Electrical World*, New York, N. Y., Sept. 30, 1933.

6. A Study of High-Voltage Flashovers, Joseph T. Lusignan, Jr. *AIEE Transactions*, vol. 48, Jan. 1929, pp. 246–54.

7. The Influence of Polarity on High-Voltage Discharges, F. O. McMillan, E. C. Starr. *Ibid.*, vol. 50, Mar. 1931, pp. 23–35.

8. Coordination of Insulation, V. M. Montsinger, W. L. Lloyd, Jr., J. E. Clem. *Ibid.*, vol. 52, June 1933, pp. 417–27.

9. La mesure des hautes tension, Schuepp, R. Van Couwenberghe, J. Kopeliovitch. *Proceedings*, CIGRE, Paris, France, Tome II, 1931, pp. 25–36.

10. Calibration of the Sphere Gap, J. R. Meador. *AIEE Transactions*, vol. 53, June 1934, pp. 942–48.

11. Impulse Calibration of Sphere Gaps, P. L. Bellaschi, P. H. McAuley. *Electric Journal*, Pittsburgh, Pa., vol. 31, June 1934, p. 228.

12. A New Correlation of Sphere-Gap Data, D. W. Ver Planck. *AIEE Transactions*, vol. 57, Jan. 1938, pp. 45–49.

13. Sixty-Cycle Calibration of the 50-Centimeter Sphere Gap, C. S. Sprague, G. Gold. *Ibid.*, vol. 56, May 1937, p. 594.

14. Effect of Ultraviolet on Breakdown Voltage, G. L. Nord. *Ibid.*, vol. 54, Sept. 1935, pp. 955–58.

15. The Irradiation of Spark Gaps for Voltage Measurement, D. R. Hardy, J. D. Craggs. *Ibid.*, vol. 69, pt. I, 1950, pp. 584–90.

16. Sphere Gap Characteristics on Very Short Impulses, P. L. Bellaschi, W. L. Teague. *Electric Journal*, Pittsburgh, Pa., Mar. 1935.

17. Short-Time Spark-Over of Gaps, J. H. Hagenguth. *AIEE Transactions*, vol. 56, Jan. 1937, pp. 67–76.

18. The Influence of Irradiation on the Measurement of Impulse Voltages with Sphere-Gaps, J. M. Meek. *Journal*, Institution of Electrical Engineers, London, England, vol. 93, pt. II, Apr. 1946, p. 97.

19. Sixty-Cycle and Impulse Sparkover of Large Gap Spacings, J. H. Hagenguth, A. F. Rohlfs, W. J. Degnan. *AIEE Transactions*, vol. 71, pt. III, Jan. 1952, pp. 455–60.

20. Polarity Limits of the Sphere Gap, F. O. McMillan. *Ibid.*, vol. 58, Feb. 1939, pp. 56–61.

21. The Dielectric Strength of Air, A. Russel. *The Philosophical Magazine*, London, England. vol. XI, Jan. 1906, pp. 237–76.

22. Dielectric Phenomena in High Voltage Engineering (book), F. W. Peek, Jr. McGraw-Hill Book Company, Inc., New York, N. Y., Sept. 1929, pp. 27–28.

23. Analog Methods in Computation and Simulation (book), W. W. Soroka. McGraw-Hill Book Company, Inc., 1954.

24. A Treatise on Electricity and Magnetism, C. A. Maxwell. Oxford University Press, London, England, vol. II, first edition, 1873; vol. II, third edition, 1892, art. 647–51.

25. Improving Field Analogues Through Conformal Mapping, H. K. Farr, W. A. Keen, Jr. *AIEE Transactions*, vol. 74, pt. I, July 1955, pp. 395–400.

26. An Electrolytic Tank for the Measurement of Steady-State Response, Transient Response and Allied Properties of Networks, A. R. Boothroyd, E. C. Cherry, R. Makar. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 96, pt. 1, no. 99, May 1949, pp. 163–77.

27. Methods of Plotting Electrostatic Fields, P. E. Green. *Engineering School Bulletin*, North Carolina State College, Raleigh, N. C. no. 45, June 1945.

28. Some Engineering Applications of the Electrolytic Field Analyzer, H. K. Farr, W. R. Wilson. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1301–09.

◆

Discussion

N. Hyltén-Cavallius (Allmänna Svenska Elektriska Aktiebolaget (ASEA), Ludvika, Sweden): This report is of great value for national as well as international standardization. When a summary of the measurements given in the paper was presented at a meeting of the IEC (International Electrotechnical Commission), we decided to investigate a few points in our own laboratory. Our tests were made with alternating voltage using 12.5- and 25-cm spheres, and were intended to show the influence of the horizontal ground plane P/D. The results showed good agreement with the measurements presented in the paper.

With reference to standardization, the results of the report under consideration are very serious. It is true that some national rules and also the IEC document *52* (1939) specify both maximum and minimum distances from the ground plane, thereby reducing this source of error, but other rules only specify a minimum value, e.g., the German VDE (Verein Deutscher Engenieur) *430*. It is also true that the present ASA (American Standards Association) Standard ($5 \leq P/D \leq 10$ for 50-cm and smaller spheres, proportionally reducing down to $P/D = 2.5$ for 200-cm spheres) is not sufficiently stringent to avoid serious errors.

It is possible, of course, to formulate recommendations so as to reduce the influence of the ground plane to acceptable values. The numerical values given in the IEC document *52* appear to be reasonable

($4 \leqq P/D \leqq 6$ for 50-cm spheres, and $3 \leqq P/D \leqq 4$ for 200-cm spheres). But even if these or similar values are accepted, the authors have demonstrated that the influence from other sources, especially live structures, is not negligible. Furthermore, there are a number of other factors which are not included in the paper and which also influence the flashover values. These are, for instance: the influence of the conductor (or conductors) connecting the high-voltage sphere to the test object; the irradiation level (especially for positive impulses); the humidity; dust; and dispersion. It should also be mentioned that the generally accepted flashover values for sphere gaps are sometimes based on measurements where the actual values of P/D, irradiation, and other factors were not reported. This means that the flashover values are more unreliable than we wish to believe.

What is to be done about all this? The best way to proceed, I think, is first to formulate some rules for the design of the spheres, the clearances to other structures, the irradiation level, etc. These rules should be sufficiently stringent to avoid serious discrepancies between measurements made with different arrangements of clearances and mounting, etc., within the range of the recommendations. They should also be sufficiently liberal in order to make international agreement possible.

One important factor to give in these recommendations is perhaps the restriction of the maximum sphere-gap spacing (S/D), as the influence from external fields and other factors is generally much larger for $S/D=1$ than for $S/D=0.4$, which is clearly indicated in the paper. When deciding upon this limit, it should be kept in mind that some theoretical considerations indicate more reliable flashover values for $S/D \leqq 0.3$, than for larger spacing. The 0.3 limit may however, from a practical point of view, be regarded as being unrealistically low. Personally, I would prefer never to go beyond $S/D=0.5$, and this has never been done in our laboratory during tests.

The next question concerns the flashover values to be inserted in the tables. As mentioned previously, the values now available may be regarded as unreliable. Therefore, the only way to proceed is to make a good and realistic choice or compromise between the values now available. These values, however, must be regarded as provisional. When realistic rules for clearances and mounting are accepted, I am sure that a number of laboratories will check the flashover values and report the results in national magazines or to CIGRE. Within a few years, therefore, it may be possible to revise the tables.

However, I am not quite convinced that the sphere gap is really a suitable measuring device in high-voltage measuring technique. A large sphere is rather costly because of the imposed requirements of free laboratory space, while both large and small spheres give rather unreliable flashover values, and will continue to do so even in the future.

It seems to me that instead of spending too much time and work on the sphere gap, it may be more economical to develop and perfect simple methods for checking the other types of measuring devices which are presently available, such as potential dividers used in conjunction with an instrument for alternating voltage, or oscillograph for impulse voltage.

It should be observed that there is no simple relation between the field factor f and the flashover voltage of a gap. This means that conclusions reached, e.g., from measurements with the electrolytic tank, are not too reliable.

With reference to the gap itself, would it be possible to publish a drawing showing the gaps, especially the arrangement of the framework?

F. M. Bruce (The Royal College of Science and Technology, Glasgow, Scotland): The data given in this paper are not only important in leading to a specification of the conditions under which a standard calibration can be applied to a sphere gap, but also in illustrating the inherent difficulties in using sparkover in a divergent field as a means of voltage measurement. The ideal spark gap would give sparkover in a uniform field, the fact that it is uniform indicating that there is no distortion due to extraneous objects. I have calibrated such gaps for 50-cycle-per-second alternating voltages up to 300-kv peak[1] and obtained data on proximity effects. Comparative data were simultaneously taken for 25-cm-diameter spheres. The results showed the remarkable stability obtained with the uniform field gaps, because of the guard-ring action of the curved edges which shield the uniform field between the central areas from distortion due to surrounding objects. We are now working on an extension of the range to several times the foregoing calibration limit, and including d-c and impulse data. Present indications are that one calibration will serve for all these conditions.

At the very high voltages now required, laboratory clearances for 200-cm-diameter spheres may well be impracticable if the gap is to be unaffected by proximity effects. An alternative might be to assess the proximity effects and apply a correction, with consequent deterioration in over-all accuracy and confidence in the results. At such high voltages, the uniform field gap has many claims to consideration.

The dimensions of extraneous objects giving rise to field distortion have a bearing on the magnitude of the effect on the sparkover voltages. Impulse generators and voltage dividers have an approximately uniform voltage gradient throughout their height, and in some cases we have found that this can materially reduce divergence of the field in a spark gap, and increase the breakdown voltage.

REFERENCE

1. CALIBRATION OF UNIFORM FIELD SPARK GAPS FOR HIGH-VOLTAGE MEASUREMENT AT POWER FREQUENCIES, F. M. Bruce. *Journal*, Institution of Electrical Engineers, London, England, vol. 94, pt. II, 1947, p. 138.

A. S. Denholm (National Research Council, Ottawa, Ont., Canada): The sphere gap is one of the most widely used methods of voltage determination in high-voltage laboratories, even although its vagaries and the necessity for care in its use have long been known. The authors have made an interesting and important addition to the subject of sphere-gap measurements, but their results confirm, to me at least, the desirability of a more accurate and consistent means of determining impulse voltage.

One possibility is the uniform field gap, which, in general, does not seem to be given the attention it is due. Perhaps the reason for this is a disinclination on the part of engineers to venture further, when they read that the Rogowski profile is determined from the field laws of Maxwell and wrongly anticipate that required profiles will be hard to determine and electrodes difficult to make. Rogowski profiles can be calculated fairly simply,[1] but a simpler profile which provides uniform field conditions is used by Bruce (see reference 1 of the discussion by Bruce) who has done the major work on the operation and calibration of such gaps. This last shape is a solid of rotation, the half profile consisting of a straight line (the electrode flat) merging into the first quarter of a sine wave which is followed by a segment of a circle. The only critical requirements are that the opposing surface should be flat and that tangency exists where the flat meets the sine curve. Gaps formed by such electrodes have a sparkover voltage given by a simple law which is accurate to about 0.2% for direct- and power-frequency voltages. Impulse calibration experiments are in an advanced stage at the Royal College of Science and Technology under Prof. Bruce. With this gap, proximity effects such as mentioned in the paper are small, and an attractive feature is that the correct operation of the gap may be checked by noting the random occurrence of sparks in the uniform field region.

We have made satisfactory uniform field electrodes without difficulty in our machine shop by using an automatic profile lathe which followed a tracing drawn in magnetic ink ten times full scale.

REFERENCE

1. GASEOUS CONDUCTORS (book), J. D. Cobine. McGraw-Hill Book Company, Inc., New York, N. Y., 1941.

Charles J. Miller, Jr. (The Ohio Brass Company, Barberton, Ohio): Those of us who are involved in the measurement of high voltages are indebted to the authors for their fine investigation into the factors which may influence the sparkover voltage of sphere gaps. Their results certainly indicate the necessity of maintaining adequate clearances in the vicinity of the sphere gap, both with respect to ground and to the side walls of the laboratory.

Of the factors investigated, the one which appears to have the greatest influence is the height above the horizontal ground plane. Since the results given indicate that a rather orderly relationship exists between sparkover voltage and the P/D and S/D ratios, it should be possible to formulate these laws as correction factors which may be applied to the sparkover voltages that are tabulated in AIEE Standard no. 4 (reference 1 of the paper). Thus, with these correction factors, it should be possible to obtain better correlation between laboratories, especially with

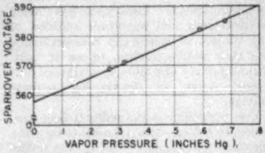


Fig. 13. Variation in sparkover voltage of 50-cm spheres as function of vapor pressure; S/D=0.6

those in which clearances may be quite limited.

In both Figs. 6 and 8 the closest correlation with AIEE sparkover voltages is obtained when the P/D ratio is in the order of 4 to 5. For a 50-cm sphere and a P/D ratio of 4, the height of the bottom of the insulated sphere would be about 6.5 feet above the floor, while for a 100-cm sphere it would be about 13 feet. It should be possible to approximate these heights above ground in most laboratories, but where this is not feasible, the proposed correction factors would provide a means for counteracting space limitations and thereby increase the accuracy of results obtained.

The effect of the distance to a vertical ground plane or the impulse generator could be minimized by selecting a suitable location of the sphere gap with respect to each. However, here again space limitations of the laboratory may not permit adequate clearances to reduce these effects to negligible values. In such cases, it may be possible for the various laboratories to obtain their own correction factors for these influences, and hence apply them to their particular setup.

H. A. Frey (General Electric Company, Baltimore, Md.): This paper presents a very interesting and valuable measurement of variations which may occur in the impulse flashover of sphere gaps due to field effects. It is particularly important that these variables be known and understood today, while consideration is being given to international and national high-voltage measurement standards.

These measurements certainly in no way discredit the sphere gap as an acceptable measuring instrument for impulse voltages, rather they merely provide information to permit the instrument to be more accurately used in such applications. Furthermore, they confirm in general the existing ASA calibrations and provide information on methods of mounting and location for maximum accuracy, as well as information on the variations which may be introduced by departures from these optimum conditions.

I believe that most laboratories today are using a dual method of high-impulse voltage measurement employing both the sphere gap and the voltage divider with cathode-ray oscillograph. This permits correlating one method of measurement against the other, and so provides a check and balance against errors in either method. This is certainly a healthy approach since it is perfectly possible that similar variables could be found in measurements made by using voltage dividers exclusively.

I should like to ask the authors if any similar measurements have been made to determine the influence of field conditions in the calibration of sphere gaps under 60-cycle voltages.

F. S. Edwards (Metropolitan Vickers Electrical Company, Ltd., Manchester, England): The data presented in this paper are at first sight very alarming as they show that errors of 18% may be obtained in the use of a sphere gap for voltage measurement if surrounding objects are not suitably arranged.

Fortunately, a closer perusal of the tables and graphs is more reassuring, but the possible errors are still too large to be ignored. The chief conclusion that emerges from the paper is the importance of restricting the ratio of spacing S to sphere diameter D. In this connection, it is interesting to note that the calibrations of F. W. Peek, Jr., (see reference 22 of the paper) extended to spacings of 1.5 diameters. However, AIEE Standard no. *4* and the British Standard no. *358*, both published in 1929, limited the spacing to about a diameter. The former document mentioned a possible error of 2%, the latter cautiously refrained from giving any guidance on this important point. In the rules accepted by the IEC in 1939, which form the basis of the current European standards, the recommended upper limit of spacing was fixed at 0.5 diameter, and the error under this condition was given as 3%. The calibrations extended to a spacing of 1.0 diameter but were stated to be less reliable from 0.5 to 0.75 diameter and to be of doubtful accuracy from 0.75 to 1.0 diameter. The 1953 edition of AIEE Standard no. *4* retained the upper limit of spacing at its old value of 1.0 diameter, with some qualifications and exceptions, but came into line with European practice by adopting 3% as the error; this figure apparently applied over the whole range of spacings.

In successive revisions, therefore, there has been a substantial reduction in the maximum recommended ratio of S/D (the corresponding reduction in the voltage range which could be measured with a given value of D was not very great, about 30%, because of the marked bending over of the voltage-spacing curve for spacings above 0.5 diameter). The accuracy claimed for the sphere gap has also diminished.

Both these changes show a more realistic attitude toward the application of the sphere gap as a means of voltage measurement, and if the present paper did no more than emphasize the futility of expecting high accuracy at large ratios of S/D without quite impracticably precise control of the disposition of surrounding objects, it would be fulfilling a very useful purpose. Actually, it does much more than that by showing that the exact value of clearances, hitherto considered to be adequate if they merely exceeded a certain minimum, may have a significant effect, and particularly that an impulse generator may seriously disturb the calibration. (It is at first sight rather paradoxical that the effect of this is greatest when the clearances to grounded objects are also greatest; the reason is obvious when once it is pointed out.) Above all, the paper gives some much needed numerical data on the errors to be expected.

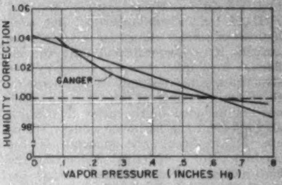


Fig. 14. Humidity correction factors

The conclusion to be reached from the results is that for reasonable accuracy the ratio S/D must be kept low, and the distance to surrounding objects must either be closely controlled or made rather greater than has previously been regarded as necessary.

The stipulation of large clearances is sometimes objected to on the grounds that the space requirements are too great to be complied with, but it is quite evident that skimping the radial clearances will invalidate the results. The choice therefore is really between a reliable measuring device taking up a lot of space and an unreliable or useless one taking up less space.

The rules for the use of sphere gaps are again being revised by a working group appointed by the IEC at its meeting in Munich in June 1956, and it is hoped that the revision will receive complete international acceptance. I am sure that those responsible for the revision will take full account of the work of the authors.

If the tables are scrutinized with the IEC rules of 1939 as a guide to the permissible values of L/D and P/D, it will be found that the grosser errors disappear completely, since the dispositions of surrounding objects which give rise to these errors are not permitted.

Thus, in Table I (Fig. 6) the departures of the flashover voltage for 50-cm spheres from the nominal value, for variations of the clearances within the prescribed limits, are ±2% for $S/D=1$, and ±1% for $S/D=0.4$.

In Table II (Fig. 8) the departures for 100-cm spheres are −0% to +3% for $S/D=1$, and −2% to 0% for $S/D=0.4$.

In Table III (Fig. 9) the departures, by interpolation, for 50-cm spheres appear to be 0% to +4% for $S/D=1$, and 0% to +3% for $S/D=0.6$.

For 50-cm spheres, the IEC rules of 1939 specify that P shall be between 8 S and 6 diameters if S is not more than 0.5 diameter, and between 4 and 6 diameters if S exceeds 0.5 diameter.

The AIEE Standard no. *4* permits P to be between 5 and 10 diameters (for 50-cm spheres), and it is easily seen from Table I how the error increases when P is increased above 6 diameters.

It is interesting to note that the authors of the paper have reached the conclusion that the solution of the problem of clearances is to reduce P/D to a value of the same order as that specified in the IEC document of 1939, thus increasing (but at the same time controlling) the influence of the ground plane, and so diminishing (relatively) the distorting effects of other objects whose positions cannot be closely specified.

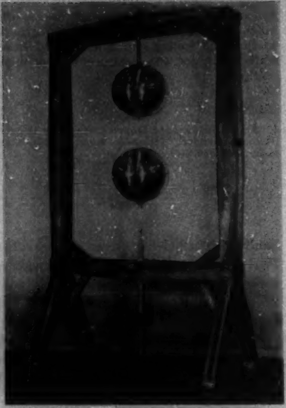


Fig. 15. Fifty-cm sphere gap

Table VI. Correlation of Critical Sparkover Voltages of 50-Cm Spheres as Related to Humidity of Air

Sparkover Crest Voltage, Kv				Diameters			Atmospheric Conditions		
Corrected for Relative Air Density	Per-Cent Variation	Corrected for Relative Air Density and Vapor Pressure	Per-Cent Variation	P/D	L/D	S/D	Relative Air Density, δ	Vapor Pressure, Inch of Mercury	Vapor Pressure Correction
699		718		4.2	14	1.0	0.983	0.18	1.028
708	1.3	720	0.28	4.2	14	1.0	0.968	0.36	1.017
726		746		6.97	14	1.0	0.983	0.18	1.028
719	0.97	738	1.08	7.02	14	1.0	0.992	0.20	1.027
795		806		14	14	1.0	0.971	0.42	1.013
794	0.1	813	0.87	14	14	1.0	0.987	0.26	1.023
569		583		4.2	14	0.6	0.992	0.27	1.023
582		583		4.2	14	0.6	0.960	0.59	1.001
585		583		4.2	14	0.6	0.955	0.68	0.995
571	2.8	583	0	4.2	14	0.6	0.969	0.32	1.020
581		596		7.02	14	0.6	0.992	0.24	1.025
593	2.05	592	0.68	7.02	14	0.6	0.960	0.66	0.997
617		616		14	14	0.6	0.960	0.66	0.997
601	2.6	616	0	14	14	0.6	0.988	0.24	1.025
584		585		4.2	9.15	0.6	0.960	0.59	1.001
574	1.74	587	0.34	4.2	9.15	0.6	0.970	0.30	1.021
602		601		7.02	9.15	0.6	0.957	0.66	0.997
592	1.7	602	0.16	7.02	9.15	0.6	0.971	0.36	1.017
622		631		14	9.15	0.6	0.965	0.39	1.015
633	1.77	632	0.16	14	9.15	0.6	0.957	0.66	0.997
577		590		4.2	4.57	0.6	0.970	0.30	1.021
581	0.69	582	1.37	4.2	4.57	0.6	0.962	0.59	1.001
597		606		7.02	4.57	0.6	0.971	0.39	1.015
608	1.84	607	0.16	7.02	4.57	0.6	0.965	0.64	0.998
641		640		14	4.57	0.6	0.957	0.66	0.997
642	0.16	653	2.0	14	4.57	0.6	0.965	0.36	1.017
450		456		4.2	14	0.4	0.964	0.43	1.012
447		458		4.2	14	0.4	0.977	0.275	1.023
444	1.35	456	0.44	4.2	14	0.4	0.985	0.19	1.028

H. E. Fiegel and W. A. Keen, Jr.: With regard to Mr. Hyltén-Cavallius' remarks concerning effects of other factors which influence the flashover values of sphere gaps, we did obtain some indications of a humidity effect during the investigation. The sparkover data presented in the paper as well as some that were not published were obtained at two different periods, 6 months apart, when the humidity was quite different. It was noted that in some cases there was as much as a 3% difference between sparkover voltages corrected for relative air density when the test conditions were identical except for rather large differences in vapor pressure. In Fig. 13 are plotted the sparkover voltages corrected for relative air density versus vapor pressure for a 50-cm sphere gap at a 0.6-diameter spacing. As will be noted, these points fall very nearly on a straight line. If a vapor pressure of 0.608 inch of mercury is used as a standard condition, a humidity correction curve as shown in Fig. 14 results. In the same figure are shown data presented by Gänger[1] which he found applied as vapor pressure corrections for uniform field gap sparkovers. A tabulation comparing sparkover data on a set of 50-cm spheres for various test conditions, when all parameters are equal except the air conditions, are presented in Table VI. Two sets of corrections are used in this table. In the first column the sparkover voltages are corrected for only relative air density whereas in the third column humidity corrections from Fig. 14 are also applied. With both relative air density and humidity corrections, 10 out of the 13 cases listed lie within a 1.0% variation, whereas the variation when corrected for relative air density only had but 4 out of 13 cases within a 1.0% variation. From these data, it appears that there is a humidity effect which should be investigated further.

Concerning his proposals for standards and revisions in standards, we agree with him in most cases and especially with his comments about using potential dividers rather than spheres, and establishing a standard method of calibrating these dividers which would be acceptable to all interested groups. In answer to the request for details of the gap arrangement, a photograph of a 50-cm sphere gap similar to the one used for the investigations is shown in Fig. 15. The 100-cm sphere gap was of the same type of construction.

We feel that the conclusions reached from the measurements with the electrolytic field analyzer were justified since we were not trying to predict the breakdown voltages of the sphere gap but rather the *change* in sparkover voltages due to the horizontal ground plane and shank-size influences. The fact that the sparkover voltages when plotted versus $(D/P)^{1/2}$ fit a straight line of the same slope as $1/f$ when plotted the same way is strong evidence of the correctness of the conclusions. This suggests that there is a simple relation between the field factor and the sparkover voltage for a given sphere diameter at a given spacing when the variable is the distance to the horizontal ground plane. This relation is

$$\text{sparkover kv} = \frac{SE_{\max}}{f}$$

where $E_{\max}$ is the critical gradient for the sphere diameter at spacing S; S is the gap spacing; and f is the field factor. The difficulty in predicting the sparkover voltage for any spacing comes about from the lack of knowledge regarding the effect of spacing on the critical breakdown gradient.

With reference to the comments of Prof. Bruce and Mr. Denholm concerning the uniform field gap as a standard for measuring high voltages, we have had no experience with the use of this gap for such purposes, but from their statements such gaps appear to be quite accurate. If however this gap were to be made a measuring standard, it would require calibration by many laboratories before being accepted. We feel that rather than going to the expense and work required to establish a new spark gap as a standard, the effort could be better spent devising and establishing standard methods of calibrating potential dividers from which the wave shape and voltage can be determined simultaneously and with one voltage application.

Dr. Miller brings up the interesting point of the possibility of formulating correction factors which could be used in conjunction with the AIEE tabulated values. This would be quite feasible except for the fact that the AIEE values were established by averaging calibrations by many laboratories with the distance to the many field-distorting influences in most cases being unknown. Thus, to use the standard values as a base to be corrected for a situation under test is not necessarily correct. We do, however, feel that knowledge of the various field-distorting influences enables the user to recognize and make proper allowances so that these influences will be minimized.

Mr. Frey points out that while the sphere gap has its faults it does turn out to be an accurate and consistent method of measuring high voltages, when used correctly. Because of its simplicity of construction and method of measurement, it provides a very good check on measurements made by a potential divider method. In answer to the question about the influences of field conditions as they affect the 60-cycle calibration of spheres, we have not made any a-c tests to determine how these field-distorting influences affect the sparkover voltage. Mr. Hyltén-Cavallius states in his discussion that he had made tests with low-frequency voltage and his results confirmed the impulse results presented in the paper.

Mr. Edwards very aptly presents a brief résumé of the evolution of how the recommended sphere spacings have been reduced as the errors in measurements made with them have been recognized. These are changes in the directions that one would expect, for the ability to measure high voltages accurately by other means has improved in this same period, and thus the error and sources of error in measurements by sphere sparkover are more readily determined. He has also discussed the information of the paper in such a way as to emphasize the salient points to various interested groups so that the sphere gap can be used more realistically for the measurement of high voltages.

REFERENCE

1. DER ELEKTRISCHE DURCHSCHLAG VON GASEN, Berthold Gänger. Julius Springer, Berlin, Germany, pp. 510, 544.

A Predicted Wave-Signalling Phase-Shift Telegraph System

E. T. HEALD
NONMEMBER AIEE

R. G. CLABAUGH
ASSOCIATE MEMBER AIEE

THE INCREASING needs of government, business, and industry have, with emphasis, pointed to the need for an advanced concept in communications. Radio and carrier spectra are at a premium. These facts, together with the knowledge that considerable improvement could be realized in system performance, provided the motivation which, in the past 8 years at Collins Radio Company[1] has resulted in the development of predicted wave signalling (PWS). PWS employs kinematic filtering and is most often used with multichannel or multiplex systems, and it has been named "Kineplex."

General

A PWS communications system accepts at its input either synchronous or asynchronous telegraph or synchronous business machine data. With conversion, the system will accept binary information from any source. The transmission system utilizes optimum detection techniques permitting improved performance in the presence of noise. The distant output, accordingly, provides signals suitable for asynchronous teleprinting or synchronous data-processing equipment.

A series of binary choices representing possible business machine or telegraph codes is transmitted wherein the following information is known by the detector before the arrival of the incoming pulse: 1. arrival and completion time, 2. pulse shape, 3. frequency and spectrum, and 4. possible data content.

Highly stable mechanical resonator filters in the detection equipment are turned on and off in step with the arrival and conclusion of each incoming signal pulse. The ability of the detector to so predict (hence, "predicted wave signalling") the incoming pulse ensures that only during intervals when the incoming signal is present will the detector be subjected to noise. It is largely this characteristic which permits superior performance in the presence of noise.[2]

The keyed detector filters are held open for T seconds and permit adjacent tone spacings as close as $1/T$ cycles.[2,3] The detector opening interval T, as represented in Fig. 1(A), is respectively for 45, 55, or 75 bit-per-second operation; 18.2, 9.1, or 9.1 milliseconds, permitting tone spacings of 110 cycles. Note that 60-, 75-, and 100-word-per-minute asynchronous telegraph speeds correspond respectively to Kineplex system speeds of 45, 55, and 75 bits per second per channel. Whether the system is employed for the transmission of binary business machine or asynchronous start-stop telegraph data, it must utilize one of these speeds.

Fig. 1(B) is a Kineplex tone frequency allocation. As shown in this figure, 20 tones are employed, each involving two channels with center frequencies ranging from 605 to 2,695 cycles; a synchronizing tone is placed at 2,915 cycles. Allowing for approximately 500 cycles of guard against unfavorable delay at band extremities, a total bandwidth of approximately 3,300 cycles is employed. This represents a two-to-one improvement in bandwidth efficiency over comparable FSK systems. Operation in bandwidths several hundred cycles narrower is possible, providing delay characteristics are favorable. Even though 20, 150 bit-per-second tone channels are each spaced 110 cycles apart, adjacent channel crosstalk can be held to approximately 40 decibels under on-frequency response.

The distant receive equipment employs a time base, the phase of which is held in step with the incoming signals by means of the transmitted 2,915-cycle synchronizing tone. This tone is on-off modulated at the transmitter so that its demodulated envelope can be phase-compared at the receiver with a similar locally generated pulse, from which phase corrections are imposed on the receiver time base to maintain receiver timing within a suitable margin.

Synchronizing tolerances of respectively ±1.9, ±4.5, and ±2.2 milliseconds are permitted for 45, 55, and 75 bit-per-second-per-channel operation. These values are the tolerable differences of delay permissible between the data tones, individually, and the synchronizing tone. The over-all delay response of the transmission medium between a Kineplex transmitter and distant receiver is generally held within these margins.

Fig. 2 is a block diagram of a duplex transmitting and receiving terminal showing only four channels. The terminal shown is equipped for 60-milliampere teleprinting inputs and outputs. Shift registers may be employed in place of code converters and keyers to make use of the entire 40 channels serially at a maximum capacity of 3,000 bits per second. In this form the system is employable for the transmission of business machine data.

Asynchronous Code Conversion

A code converter effects conversion of the 60-, 75-, and 100-word-per-minute start-stop telegraph code to a 7.0 code compatible with the Kineplex synchronous transmission system. The code converter is a single-level storage and direct read-out device and has built-in

---

Paper 57-65, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21-25, 1957. Manuscript submitted October 24, 1956; made available for printing December 17, 1956.

E. T. HEALD and R. G. CLABAUGH are with the Collins Radio Company, Burbank, Calif.

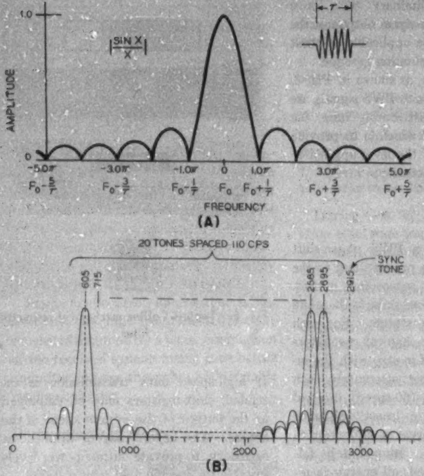


Fig. 1 (left). (A)—Frequency response of keyed resonator filter. (B)—Tone frequency allocation

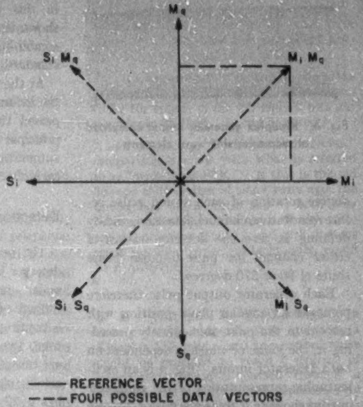

M_q
$S_i\ M_q$
$M_i\ M_q$
S_i
M_i
$S_i\ S_q$
$M_i\ S_q$
S_q
REFERENCE VECTOR
FOUR POSSIBLE DATA VECTORS

Fig. 3 (above). Vector positions representing phase shifts between successive transmitted elements

protection against false starts, the durations of which are defined as an interval of less than one quarter of one bit.

Phase-Shift Generation

Two channels of binary information are associated with each tone generator. For Kineplex telegraph employment therefore two code converters or shift-register outputs are common to each tone generator, as shown in Fig. 2. The mark-space outputs of both code converters, or the associated shift register, are transformed at the generator so that the phase shift between successive tone generator output pulses is either zero or 180 degrees for a channel defined as the "in-phase" channel, and 90 or 270 degrees for the "quadrature" channel. Considering four possible combinations of two code converter outputs, the resultant phase shifts between successive generated pulses are therefore 45, 135, 225, and 315 degrees, as indicating in Fig. 3.[4]

At the detector, both channels are rendered mutually independent by a 90-

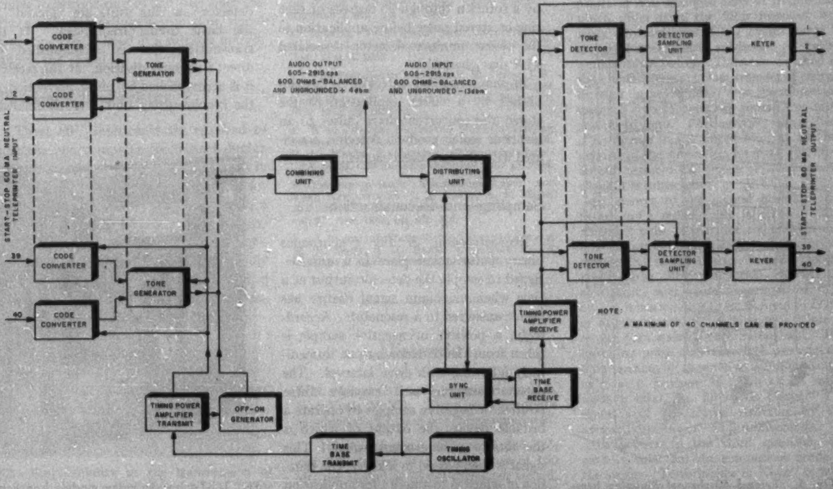

1
2
39
40
CODE CONVERTER
CODE CONVERTER
TONE GENERATOR
CODE CONVERTER
CODE CONVERTER
TONE GENERATOR
START-STOP 60 MA NEUTRAL TELEPRINTER INPUT
AUDIO OUTPUT
605-2915 cps
600 OHMS-BALANCED
AND UNGROUNDED + 4dbm
AUDIO INPUT
605-2915 cps
600 OHMS-BALANCED
AND UNGROUNDED -13dbm
COMBINING UNIT
DISTRIBUTING UNIT
TONE DETECTOR
DETECTOR SAMPLING UNIT
KEYER
TONE DETECTOR
DETECTOR SAMPLING UNIT
KEYER
START-STOP 60 MA NEUTRAL TELEPRINTER OUTPUT
TIMING POWER AMPLIFIER RECEIVE
NOTE:
A MAXIMUM OF 40 CHANNELS CAN BE PROVIDED
TIMING POWER AMPLIFIER TRANSMIT
OFF-ON GENERATOR
SYNC UNIT
TIME BASE RECEIVE
TIME BASE TRANSMIT
TIMING OSCILLATOR

Fig. 2. Block diagram of Kineplex duplex transmitter receiver telegraph terminal

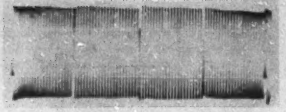


Fig. 4. Kineplex generator output comprised of successive sine wave elements

degree rotation of each stored pulse before quadrature channel detection, and by defining as zero the detector output of either channel for pulse-to-pulse phase shifts of 90 or 270 degrees.

Each generator output pulse therefore possesses an angular phase position with respect to the pulse immediately preceding it, the value of which is dependent on both generator inputs. Fig. 4 is an oscillographic representation of the generator output showing it to be a series of discrete bit-length sine waves, the frequency of which is accurately determined by a highly stable mechanical resonator oscillator.

Common Equipment

The common equipment is comprised of the combining and distributing units. As shown in Fig. 2, a combining unit accepts the outputs of the 20-tone generators and serves to combine and amplify with a minimum of distortion. Side-band components of the generated PWS signal are limited to approximately 3,400 cycles in this unit. Preliminary tests have shown the composite signal output of the combining unit to be applicable to most standard telephone circuits.

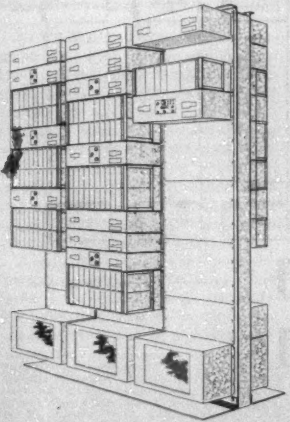


Fig. 5. Typical 40-channel Kineplex transmitter receiver terminal

At the destination, as shown in Fig. 2, the incoming composite PWS signals are passed through a distributing unit, the principal function of which is to provide automatic control of the level applied to each of 20 PWS detectors.

Detection

The output from a PWS phase-shift detector is defined as proportional to the scalar product of succeeding pulses. Storage of each previous pulse is therefore a requirement of the system. For each tone, two high-Q mechanical resonators are turned on and off in step with the incoming signal such that, alternately, each has a 22-millisecond (60-speed) interval where it may "ring" or "coast" following every "drive" or interval of subjection to the incoming pulses. Immediately following this ring period, all energy is removed from the resonator by a quenching operation so that the resonator will possess no knowledge of previous pulses. A phase comparison of the ring interval of one resonator and the concurrent drive of the other results in a detector output, the sign of which indicates the nature (mark or space) of the information represented by the phase relation between the stored and drive pulses.

As previously indicated, detection of the quadrature channel is accomplished by a rotation through 90 degrees of each ring or stored pulse before application to the scalar product detector associated with that channel.

In-phase channel detection is accomplished by a direct application of the stored and concurrent drive pulses to an identical scalar product detector associated with the in-phase channel.

Sampling and Reconstruction

Reconstruction of the synchronous binary pulses takes place in a unit designed to sample the detector output at a time when maximum signal energy has been transferred to a resonator. Accordingly, a positive or negative sample is taken from the detector output immediately following each drive interval. The detector-sampling unit employs these positive or negative samples to operate a bistable device, the output of which is the totally reconstructed signal. This signal is then passed to a keyer, the function of which is to amplify for application to 60-milliampere teleprinting circuits.

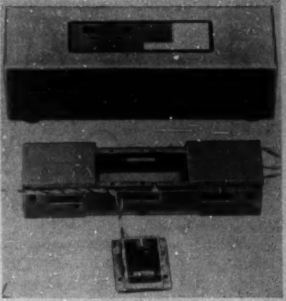


Fig. 6. Typical Collins mechanical resonator filter

If high-speed data transmission is intended, shift-registers may be employed at the output of the system ahead of the keyers. The keyers may be applied, as necessary, to provide output power levels suitable for specific load requirements.

Timing

Channel transmission rates of 45, 55, and 75 bits per second are derived from basic timing frequencies of respectively 727.3, 888.9, and 1187.2 cycles per second, which are generated in timing oscillators located at the transmitter and receiver. Timing wave forms required by the code converters or shift-registers, synchronizing equipment, detectors, and detector-sampling units are "locked" to the basic timing frequencies. At the transmitter, this is accomplished by direct frequency division; at the receiver it is accomplished by phase correction at the synchronizing unit (Fig. 2) followed

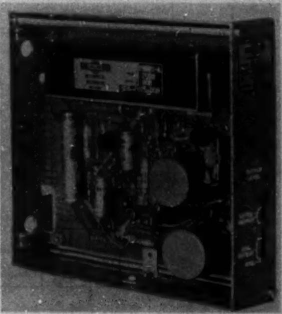


Fig. 7. Typical Kineplex module and transistorized printed circuitry

by direct receiver time base frequency division. Phase correction, it will be recalled, is necessary to hold the receiver in step with the incoming PWS signals.

Frequency Stability

Frequency errors resulting from instabilities in carrier service between Kineplex terminals may be as great as ±10 cycles and may occur at rates as fast as 1/2 cycle per second when an automatic frequency control is employed at the receiver input. Output from the automatic frequency control is held to within 1/10 cycle.

Frequency control information is derived by this unit from the 2,915-cycle synchronizing tone by a phase comparison of the received synchronizing tone before and after passage through a highly stable mechanical filter.

Performance in Noise

Laboratory tests have indicated that for a 1/10 of one per cent character error rate, a signal-to-noise ratio of −3.0 decibels in a 3-kc band is representative. This figure betters wide-band FSK performance by approximately 8 decibels.[5,6]

Protection against impulse noise is expressly built in for each detector. Limiting is automatically imposed on instantaneous noise bursts, the amplitudes of which are greater than 13 decibels over the composite signal level. Long-time increases in signal level cause no appreciable distortion due to limiting, since the threshold of limiting is, for slow signal variations, held fixed at a suitable margin above the signal.

Physical Characteristics

Three $86^{1}/_{2}$-inch racks are required to contain one complete 40-channel duplex transmitter-receiver terminal, as shown in Fig. 5. The total power requirement for such a terminal is approximately 1 kw; a total of over 2,000 transistors and approximately 65 mechanical resonators are employed. Figs. 6 and 7 show respectively a typical mechanical resonator and a typical module. Fig. 7 typifies the modular construction and printed circuitry.

Summary

The Kineplex system is characterized by the following:

1. High capacity, high bandwidth efficiency.
2. Employability in the transmission of either business machine data or 60-, 70- or 100-word-per-minute commercial telegraph.
3. Minimum crosstalk between adjacent channels and between successive signal elements.
4. Superior performance in the presence of noise.
5. Transistorized printed circuitry and attendant reliability.
6. Small physical size and power requirement.
7. Direct applicability to commercial wire and carrier transmission services.

During October 1956, Kineplex evaluation tests were conducted over telephone wire and carrier circuits selected by the Bell Telephone Laboratories. A serial 450 bit-per-second system was tested over carrier and microwave facilities originating and terminating in the Burbank, Calif., laboratories of the Collins Radio Company, with relay points located at such distant locations as Santa Barbara and San Bernardino, Calif.; Las Vegas, Nev.; and Denver, Colo.

A Kineplex system, involving four 60-word-per-minute channels with loading equivalent to 40 channels, has been under test for an extended period over a high-frequency radio path between Cedar Rapids, Iowa and Canoga Park, Calif., a distance of 1,800 miles. The results of these telephone and radio frequency tests have confirmed conclusions regarding Kineplex system characteristics relating to operation over both media.

References

1. Radiotelegraphy, W. H. Wirkler. *Engineering Report CER-W241*, Collins Radio Company, Burbank, Calif., Apr. 12, 1949.

2. Special Techniques for Detection in Noise, M. L. Doelz. *Engineering Report CER-W272, Ibid.*, Mar. 1, 1953.

3. The Philosophy of PCM, B. M. Oliver, J. R. Pierce, C. E. Shannon. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 36, no. 11, Nov. 1948, pp. 1324–31.

4. A 40 Channel Predicted Wave Multiplex System Using Phase-Shift Keying, E. T. Heald, D. L. Martin, F. Secretan. *Engineering Report CER-W356*, Collins Radio Company, July 15, 1954.

5. Frequency-Shift Telegraphy-Radio and Wire Applications, J. R. Davey, A. L. Matte. *Bell System Technical Journal*, New York, N. Y., vol. 27, Apr. 1948, pp. 265–304.

6. A Predicted Wave Radio Teletype System, M. L. Doelz, E. T. Heald. *Convention Record*, Institute of Radio Engineers, pt. 8, 1954.

◆

Discussion

J. B. Moore (RCA Communications, Inc., New York, N. Y.): There are a few items on which clarification would be appreciated:

1. In effecting phase shifts of 45, 135, 225, and 315 degrees at a keying rate of 150 bits per second or 75 dot-cycles per second, in a total bandwidth of perhaps 1/2 to 2/3× 110 cps (cycles per second), say, 50 to 67 cps total, is it possible for the new phase to reach its steady-state value within the time allotted? If not, how does this affect keying fidelity and signal-noise ratios?

2. Synchronizing tolerances are stated as ±1.9, ±4.5, and ±2.2 milliseconds respectively for speeds of 45, 55, and 75 bits per second per channel. Is ±4.5 correct?

3. In the section "Performance in Noise," comparison is made with what is referred to as "wide-band FSK." If this is the old 850-cps shift system of some years ago, the comparison means little and is both misleading and unfair. Comparison with commercial practices of today, and with systems made possible by present techniques, would be a different story.

E. T. Heald and **R. G. Clabaugh:** In answer to Mr. Moore's first question, a Kineplex 13.5-millisecond transmitted pulse might be expected to reach 5 degrees of steady-state phase if passed through a passive-tuned circuit filter of approximately a 50-cycle bandwidth. This question is interpreted to concern the manner in which the ensemble of 20 Kineplex data tones are each placed 110 cycles apart and separated at the receiver with negligible crosstalk.

The question is particularly worth while because it points out a significant feature of the Kineplex system. Narrow passive filtering is not a part of Kineplex design. The Kineplex optimum filtering technique employs an active or time varient Kinematic filter, the response of which is essentially identical to the spectrum of the signal to be filtered. It can be shown that maximum discrimination between signal and noise generally results from a filtering process optimum in the sense just mentioned.

Figure 1(A) of the paper represents the response of a keyed or time variant Kineplex filter. It is to be observed that nulls or points of zero response exist at frequencies which differ from the filter resonant frequency by multiples of $1/T$ cycles where T is the interval of filter subjection to the incoming signal pulses. Data tones placed at these nulls produce no net transfer of energy through the keyed filter; over the subjection interval T these tones are summed or integrated to zero. Furthermore, since there is no energy in the keyed filter at the beginning of the subjection interval, the filter output assumes the phase of the pulse to which it is subjected.

In answer to the second question, the figure ±4.5 milliseconds is correct. The Kineplex pulse length, for a channel transmission rate of 55 bits per second, is approximately 18.2 milliseconds. To maintain the predetection filter subjection interval T, consistent with tone spacings of 110 cycles, it is necessary to set T equal to 1/110 or approximately 9.1 milliseconds (time is required to quench each resonator). The incoming pulse may therefore precede the filter opening by as much as 9.1 milliseconds or it may conclude as much as 9.1 milliseconds after the filter has closed. The signal may be delayed therefore by the nominal amount of ±4.5 milliseconds.

In answer to the third question, the authors hold that the comparison made in the section "Performance in Noise" is both standard and well-referenced. A comparison with any specific system appears to be a matter of choice.

Economics of Multimillion-Joule Inductive Energy Storage

H. C. EARLY
MEMBER AIEE

R. C. WALKER
MEMBER AIEE

IN MANY electric energy storage applications the objective is to store energy slowly over a relatively long interval and than withdraw the energy quickly to obtain a very high peak power. For certain applications of this type, inductance coils have significant advantages over capacitors or rotating machinery. The inductive storage method is particularly advantageous in situations where millions of watt-seconds are to be stored and the energy is to be expended in a time interval of a few milliseconds.

This paper is based on the experience gained in building an inductance coil for storing 3 megajoules of energy, which is used to produce a high-power arc at high gas pressure. The design and performance of the present installation is described, and some of the merits and limitations of other coil designs and alternative energy storage systems are also considered.

The point of view of the discussion is based on the assumption of a low duty-cycle operation where cumulative heating due to repetitive operations can be ignored, and where the only limiting factors are the instantaneous ratings of equipment. This assumption results in a very large reduction in the size and cost of the equipment.

Comparison of Storage Methods

The four most useful electric devices for storing large quantities of energy appear to be capacitors, inductances, rotating machinery, and storage batteries. For certain applications the choice among competing methods can be made primarily on the basis of cost, but in most situations factors other than the cost per watt-second are at least of equal importance.

Probably the most important differentiating characteristic among these systems is the length of the discharge time. When the discharge interval must be short, compared to a millisecond, capacitors have the advantage. For cases in which the discharge interval is of the order of milliseconds, the inductance coil is particularly attractive. For slower rates of discharge, corresponding to an interval of perhaps a second, rotating machinery is advantageous, while for still longer discharge periods (and long storage periods), storage batteries have an unusual advantage. In brief, as a rough rule of thumb, the four systems might be characterized by their possible minimum discharge periods: capacitor, microseconds; inductance, milliseconds; machinery, seconds; battery, minutes.

While capacitors have a unique advantage in their short discharge time, they have a serious economic limitation in comparison with inductance coils, machinery, and batteries, in that they cannot be temporarily overloaded regardless of how low the duty cycle may be. The cost of a capacitor bank varies linearly with the amount of energy to be stored. At present, energy storage capacitors cost about 6 cents per watt-second, and their volume is approximately 1 cubic inch per watt-second. The cost of the complete installation is much greater than these figures would indicate because of space requirements, shelf-type mounting arrangements, and wiring which requires an extensive protective system of relays and interlocks.

In comparing capacitive versus inductive storage, the differences in circuit behavior of the two systems may be a deciding factor. For instance, when a capacitor bank is discharged through a high-pressure electric arc, the arc will often extinguish when the capacitors are only partially discharged. On the other hand, if the arc energy is supplied from an inductance coil, the arc cannot extinguish until all the stored energy has been expended. In general, the capacitor tends to act as a voltage generator and the current is determined by the load characteristics, while the inductance tends to act as a current generator and the voltage is determined by the load characteristics.

A University of Michigan research project required the storage of 1 to 3 million watt-seconds of energy, to be dissipated in an arc discharge in a period of about 10 milliseconds. The short discharge time ruled out the possibility of using machinery or batteries, while the linear increase in cost of capacitors made their use very expensive. Therefore, the possible use of a coil was examined. At the million-joule level, it was found that an inductive storage system would be cheaper than the capacitive system. However, a more important conclusion is that the inductive storage installation becomes very much cheaper, in terms of watt-seconds per dollar, as the size increases. This occurs primarily because of a scaling factor whereby the ratio of stored energy to peak charging power of an inductance increases with the square of a linear dimension of the coil. Hence, the charging period for a larger coil can be extended over a longer interval, and the peak power requirement from the d-c supply is greatly reduced.

Specifications of University of Michigan Coil

Fig. 1 shows the inductance which is designed for storing 1 million joules when operated as an air-core inductance, and 3 million joules when the center is filled with transformer iron. The wood reel is 10 feet in diameter and 3 feet wide, and the inside core or drum is 5 feet in diameter. The winding consists of 5,000 feet of 3 million-circular-mil 169-strand aluminum cable; the total weight of the conductor is 15,000 pounds. There are $12^1/_2$ layers on the winding and a total of 220 turns. The insulation for the cable consists of two half-lapped wrappings of 5-mil Mylar tape covered with two wrappings of 8-mil adhesive-paper tape. The interlayer insulation consists of Mylar sheets for electrical insulation and canvas for mechanical-stress distribution.

Without iron in the core, the coil has an inductance of 75 millihenrys, a time constant of 2.5 seconds, and self-resonant frequency of 16 kc. A current of approximately 5,200 amperes is required to store 1 megajoule. The d-c resistance is 0.03 ohm, and 5,200 amperes of current requires 800 kw of d-c power which is furnished by a 3-phase ignitron rectifier. The current builds up at an exponential rate during the charging period which lasts approximately 6 seconds.

Economics of Scaling Factors

An economical design of energy storage apparatus requires a proper balance between the cost of the coil and the cost of

---

Paper 57-79, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted March 19, 1956; made available for printing December 17, 1956.

H. C. EARLY and R. C. WALKER are with the University of Michigan, Ann Arbor, Mich.

Fig. 1. Energy storage coil

the power supply. For a given storage capacity, the more money that is spent on the coil, the less is required for the power supply and vice versa. The manner in which these relative costs and also the total costs vary with the capacity of the installation is considered in this section. The discussion is based on several simplifying assumptions which appear to be generally true. However, in certain special cases, they might require reviewing. These assumptions are:

1. The duty cycle is very low and the only limit to the instantaneous current density in the winding is the economic factor involving the cost of the power supply. (Perhaps the only situation where this assumption breaks down seriously is when the discharge time is unusually short and the peak power and voltage are so large that insulation breakdown becomes a serious factor.)
2. The cost of the d-c power supply is proportional to the peak capacity in kilowatts. (This assumption holds roughly over a considerable range, but at sufficiently large power levels the cost per kilowatt tends to become less.)
3. The cost of the coil is proportional to its volume, i.e., to the cube of a linear dimension.

In evaluating various coil designs, a figure of merit may be defined which relates the total energy stored at peak current to the peak charging power.

$$\text{Figure of merit} = \frac{\text{stored energy}}{\text{peak charging power}} = \frac{1/2LI_m^2}{I_m^2R} = 1/2\frac{L}{R} = 1/2\ (\text{coil time constant})$$

where

L = inductance of the coil, henrys
R = resistance of the coil, ohms
I_m = peak coil current, amperes

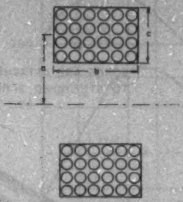


Fig. 2. Multiple-layer coil dimensions

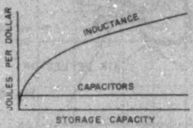


Fig. 3. Trend of joules per dollar versus energy storage capacity

This definition of a figure of merit is helpful in comparing the relative advantages of different coil sizes and shapes. Consider a multiple-layer coil, having a mean radius a, as shown in Fig. 2. For a given coil shape, the inductance is proportional to the square of the number of turns and to the first power of a given coil dimension, $L=(\text{constant})\ an.^2$ The resistance is also proportional to the square of the number of turns, but is inversely proportional to the first power of a given coil dimension, $R=(\text{constant})\ n^2/a.$[1,2]

Therefore, for any given shape, the following obtains

$$\text{figure of merit} = 1/2\frac{L}{R} = 1/2\frac{(\text{constant})an^2}{(\text{constant})n^2/a} = (\text{constant})a^2$$

and the relation of stored energy to peak charging power varies as the square of the given coil dimension. Thus, for a given volume or weight of conductor, the figure of merit is immediately specified regardless of whether the coil is wound of many turns of fine wire or few turns of large wire. (This statement neglects the effect on L and R of insulation spacing.) From the foregoing relationship, it can be seen that increasing the linear dimensions of a coil by a factor k will increase the ratio of stored energy to peak power by a k^2 factor.

On the basis of the assumptions stated previously, several important relations can be derived.

Let

$\$_p$ = cost of power supply
$\$_c$ = cost of coil

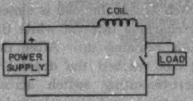


Fig. 4. Basic circuit for inductive energy storage

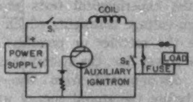


Fig. 5. Modified circuit

Then

$$\text{figure of merit} = \frac{\text{joules}}{\text{watt}} = (\text{constant})\ (\text{coil volume})^{2/3} = (\text{constant})\ (\$_c)^{2/3}$$

$$\text{storage capacity} = \frac{\text{joules}}{\text{watt}}\ (\text{peak power}) = (\text{constant})(\$_c)^{2/3}(\$_p)$$

$$\text{joules/dollar} = \frac{\text{storage capacity}}{\text{total cost}} = (\text{constant})\frac{(\$_c)^{2/3}(\$_p)}{(\$_c+\$_p)}$$

For a given total cost $(\$_c+\$_p)$, maximum stored energy is obtained for an optimum ratio $\$_p/\$_c$. The optimum ratio is 3/2. However, the ratio $\$_p/\$_c$ can vary from unity to 5/2 without decreasing the joules/dollar ratio more than 4%.

By substituting $\$_p/\$_c=3/2$ in the foregoing relations, the following results are obtained:

$\text{total cost} = \$_c+\$_p = 5/2\ \$_c = 5/3\ \$_p$
$\text{storage capacity} = \text{constant}\ (\$_c)^{5/3} = \text{constant}\ (\text{total cost})^{5/3}$
$\text{joules/dollar} = \text{constant}\ (\text{total cost})^{2/3}$
$\text{joules/dollar} = \text{constant}\ (\text{storage capacity})^{2/5}$

Fig. 3 illustrates the trend of joules/dollar of capacitive and inductive storage, as storage capacity of a system is increased. For the small installation, capacitors are less expensive, but the inductance becomes progressively more advantageous as the size of the system increases. The location of the crossover point depends on many factors which cannot be expressed analytically. However, based on present experience, the crossover point is probably at energy levels of the order of 100,000 watt-seconds.

Circuit Switching

The simplest possible circuit-switching arrangement for charging and discharging the coil is shown in Fig. 4.

The switch across the load is closed during the charging interval while the coil current is building up. After the coil current has reached the desired value, the short-circuiting switch across the load is opened, and the relatively high resistance of the load is inserted into the circuit. Two modifications of this basic circuit were found to be desirable: 1. Since the experimental tests produced fast high-voltage transients which might damage the rectifier or possibly reflect into the power line, the system was completely disconnected from the power source before the stored energy was delivered to the load. 2. To open the switch across the load without arcing, an auxiliary by-pass circuit, described in the following, was connected across the switch terminals.

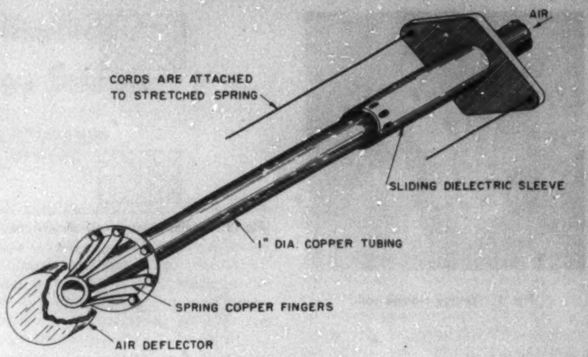

AIR
CORDS ARE ATTACHED TO STRETCHED SPRING
SLIDING DIELECTRIC SLEEVE
1" DIA. COPPER TUBING
SPRING COPPER FINGERS
AIR DEFLECTOR

Fig. 6. Fast-acting switch

The coil and discharge circuit are disconnected from the rectifier, before starting the discharge through the load, by means of an auxiliary ignitron, as shown in Fig. 5. At the completion of the charging period, opening switch S_1 disconnects the rectifier. The voltage induced by the coil fires the auxiliary ignitron and the current is switched from the rectifier to the discharge ignitron. The switch S_1 is not required to handle any high voltages, since the maximum voltage is limited to the d-c rectifier voltage which is about 200 volts.

The switching operation at S_2 (Fig. 5), whereby the coil current is switched into the load, is more difficult. Because the current is switched into a high-impedance load, voltages of tens of thousands of volts are developed across the switch. This switching problem was solved by constructing an unusually fast opening switch, which was used in conjunction with a high-voltage fuse, as shown in Fig. 5. The operation is as follows:

1. When the switch S_2 opens, the current is first transferred to the high-voltage fuse. Because of this short-circuiting fuse, there is no arcing as the switch opens. This allows the switch several milliseconds to develop a high breakdown strength before the fuse blows.

2. After about 5 milliseconds the fuse blows and the current is delivered to the load. The spark gap in series with the load prevents any load current from flowing until the voltage has reached a value of about 50 kv.

The fast-acting mechanical switch S_2 is shown in Fig. 6. One terminal of the switch is a 1-inch-diameter copper tube, and the other terminal is a circular ring of spring-loaded fingers which make contact with one end of the pipe when the switch is closed. A thin cylindrical sleeve covered with Mylar insulation slides freely along the pipe and opens the circuit when it passes between the spring-loaded fingers and the tube. The dielectric sleeve is accelerated by means of a stretched rubber spring. The sleeve has a small mass, and is accelerated to a velocity of 200 feet per second in 2 feet of travel. This velocity is more than adequate for present requirements. To minimize any arc which might tend to form across the opening switch, a stream of compressed air is used to blast the dielectric surfaces along which an arc might form. This air-blast arrangement is shown in the figure.

The expendable high-voltage fuse consists of an 18-inch length of no. 18 copper wire immersed in an oil-filled fiber tube. Experience with this switching arrangement indicates that it could readily be adapted to handle much higher voltages and currents and faster switching rates than have been required by the present installation.

Power Sources

For power-supply voltages of more than 200 volts, the ignitron rectifier is efficient and economical. For somewhat lower voltages, selenium, silicon, and germanium rectifiers are more feasible, while for extremely low potentials of only a few volts, the homopolar generator is particularly attractive.

If the peak power requirement of the energy storage coil is greater than that which can be supplied by the available power-line capacity, an intermediate storage system, e.g., a flywheel, can be used. Energy is taken from the power lines over a relatively long time period, perhaps a minute, to be stored in the flywheel. This kinetic energy is transferred to the coil in seconds, and then discharged to the load in milliseconds. An example of what can be done with flywheel energy storage is the bevatron installation at the University of California. Two 67-ton flywheels store 675 million watt-seconds at synchronous speed. Over a period of 1.7 seconds, 80 million joules are taken from the flywheels. The speed change of the flywheel is about 6% during this discharge period.

Impedance Matching

When the coil is discharged into a load of constant resistance, the current decreases in an exponential manner during the discharge interval. The power during this interval varies as the square of the current and thus also decreases at an exponential rate. Other power-pulse shapes are produced when the load is a nonlinear resistance, such as an arc. A number of methods for varying the arc resistance during the discharge interval have been tested, and a variety of power-pulse shapes have been obtained.

Tests with the coil of Fig. 1 have involved arc impedances of 5 to 20 ohms. For many applications of a stored energy system, operation with much lower load impedances and voltages is desirable. This can be accomplished by using a coil with fewer turns and larger currents. However, charging a very low-impedance coil requires very large currents at such low voltages that the cost of the d-c power supply becomes more significant.

An alternate method of operation with a very low impedance load is being investigated. The energy storage inductance coil is being rebuilt, and a low-impedance secondary winding is being added so that much higher discharge currents can be obtained. The transformer action provided by a secondary winding offers some attractive advantages. In the design of an installation for delivering currents of the order of 1 million amperes, these advantages become quite significant:

1. The primary circuit operates at a higher voltage and lower current, and the requirements for the d-c power are simplified.
2. The switching operation for charging and discharging the coil can be handled in the primary circuit and, hence, equipment for switching currents of the order of 1 million amperes in a few milliseconds of time is not required.
3. The secondary winding is isolated from the primary, and therefore the load circuit can be grounded at any desired point.

The effective use of a secondary winding requires close coupling between the two windings. Otherwise, an excessive amount of energy is lost in switching the current from the primary to the secondary. A number of small-scale models of transformer-type coils with interleaved windings have been built and tested, and coupling coefficients greater than 0.95 can be obtained.

Multiple-Layer Proximity Effects

If most of the energy stored in a coil is to be delivered to an external load, the load resistance must be much greater than the coil resistance. If the coil has a low Q, i.e., a high effective a-c resistance, for the important components of the discharge-pulse frequency spectrum, the energy dissipated in the coil will be very much greater than that predicted by a simple comparison of the coil d-c resistance to the load resistance.

A multiple-layer coil tends to have a much higher ratio of a-c resistance to d-c resistance than a single-layer coil. The increase in a-c resistance due to proximity effect is many times larger than would be predicted by the formulas for skin effect in an isolated conductor. In the multiple-layer coil, each turn is located in a magnetic field produced by all of the surrounding turns, which induces complex eddy currents within the conductors. For example, consider the coil in Fig. 1. If the effect of stranding is neglected, and the cable considered as a solid conductor, at 60 cycles the R_{ac}/R_{dc} ratio is about 100, due to proximity effect. If the same cable were used as a conductor on a power line, the R_{ac}/R_{dc} ratio due to ordinary skin effect would be 1.15.

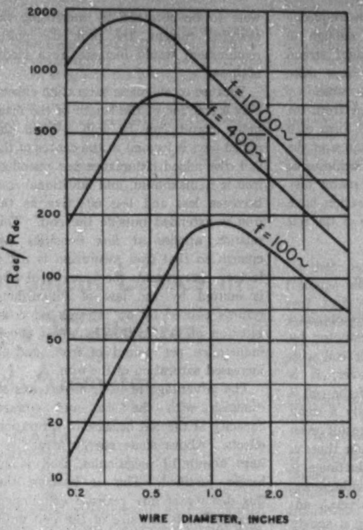


Fig. 7. Variation of R_{ac}/R_{dc} of multiple-layer coil of Fig. 1 as a function of frequency and wire diameter

Butterworth[3] has published a comprehensive investigation of coil losses due to skin and proximity effects. His formulas are developed for two types of conductors: 1. solid round wire, and 2. litz cable, in which the individual strands are insulated from each other. The ratio of a-c to d-c resistance is expressed as a function of both frequency and wire diameter. (The wire diameter is the actual conductor diameter for solid wires; for litz wire, the wire diameter is the individual strand diameter.) Calculated curves for various frequencies and diameters for a coil of the over-all dimensions of Fig. 1 are shown in Fig. 7. These calculated curves were checked experimentally by means of models and the use of the scaling laws given in reference 2 of the paper.

A coil wound of litz cable often has a R_{ac}/R_{dc} ratio which is smaller, by at least an order of magnitude, than the ratio for a similar coil wound of solid conductor. The intermediate case of a stranded conductor, in which there is no interstrand insulation, lies between these two extremes and is difficult to evaluate. Because of the contact resistance between the strands, the stranded conductor tends to have a partial "litzing" effect and in many cases has a lower R_{ac}/R_{dc} ratio than the equivalent solid conductor. This effect appears to be more marked in aluminum stranded cables than in copper stranded cables. The contact resistance is nonlinear, since it depends on the characteristics of oxide film between the strands and it is difficult to measure or estimate. Tests at the University of Michigan on a 1 million-circular-mil 61-strand aluminum cable indicated that this interstrand resistance was several times greater than that of similar copper cable. A recently published paper,[4] describing tests made under power-line conditions, reports lower R_{ac}/R_{dc} ratios for aluminum than for copper. These tests were made at 60 cycles in stranded cables of the 2,250,000-circular-mil size.

The coil shown in Fig. 1 is wound with 3 million-circular-mil stranded cable with no insulation between the strands. Fig. 7 gives the resistance ratio which would be expected at various frequencies if the interstrand resistance were assumed to be zero, i.e., a solid 3 million-circular-mil cable. For a discharge time of 1 millisecond, the important frequency components are high enough so that the a-c resistance would be excessive, but for the discharge time of 10 milliseconds it was possible to use ordinary stranded aluminum cable with uninsulated strands. For this discharge time, only the low-frequency components of the discharge-pulse frequency spectrum (perhaps up to 25 cycles) have sufficient magnitude to be important. The increase in effective coil resistance due to these low-frequency components was not great enough to warrant the cost of individual strand enameling, since this would have more than doubled the cost of the cable.

The 10-millisecond discharge from the 75-millihenry coil was in the form of a power arc which had a resistance of the order of 5 to 20 ohms. Oscilloscopic measurements made during many discharge tests of this time duration have shown that the a-c resistance of the coil did not exceed 1 ohm.

Coil Geometry and Winding

Maximum inductance can be obtained with a given length of conductor by winding it as a multiple-layer coil with $b=c=0.66a$ (Fig. 2). However, it is important to point out that if the coil is wound as a long solenoid, only a small decrease in inductance will result even when b becomes as much as four times c. At $b=4c=1.2a$, the coil inductance is still 88% of the maximum inductance.

The long solenoid has several advantages over the coil having a square cross section. Since the number of layers in the coil is reduced, many winding problems in the coil construction are minimized. Also, during the charging and discharging of the coil, the magnetic forces acting on the turns tend to scuff and abraid the insulation. By reducing the number of layers, the load that is supported on the inner layers of the coil is reduced so that there is less wear on the insulation of these turns. From an electrical standpoint, the increase in a-c resistance of the coil during discharge due to proximity effect is greatly reduced, since proximity effect is proportional to the square of the number of layers.

Use of Iron in the Magnetic Circuit

The University of Michigan coil has been used so far without any iron in the magnetic circuit. For 1 million joules of energy storage, the flux density in the center of the air coil is 5,000 gauss and the peak d-c charging power is 800 kw. The planned addition of 30,000 pounds of transformer iron into the hollow core of the wooden coil form will increase the inductance by a factor of three, and the stored energy, for the same peak charging power, will be correspondingly increased without saturating the iron. This calculation has been confirmed by scale-model tests. With this coil, the cost of the iron is economically justified because if the same increase in storage capacity were to be obtained by increasing the charging current, the peak d-c power requirement would increase by a factor of three.

Because of magnetic saturation effects, only a fraction of the length of the magnetic circuit can be iron. When the added iron is located in the center of the coil, the added inductance per pound of iron is a maximum, and additional iron becomes less and less effective as the iron is extended outside the coil. This relation applies at flux densities low enough so that iron saturation is not a factor. In general, the amount of iron is limited by the law of diminishing returns caused by an interrelated combination of two factors: 1. less added inductance per pound of iron, and 2. increased saturation of the iron.

The advantage of using iron tends to diminish with the size and storage capacity of the coil because of saturation effects. Above some energy level, perhaps about 10 megajoules, iron is no longer justified. This is because the flux density of any realistic coil design increases with the size of the coil until the use of iron can no longer be justified economically.

References

1. Radio Engineers' Handbook, **F. E. Terman.** McGraw-Hill Book Company, Inc., New York, N. Y., 1943, p. 61.

2. Design of Standards of Inductance, and the Proposed Use of Model Reactors in the Design of Air-Core and Iron-Core Reactors, **H. B. Brooks.** *Journal of Research*, U. S. Bureau of Standards, Washington, D. C., vol. 7, Aug. 1931, p. 289.

3. Effective Resistance of Inductance Coils, **S. Butterworth.** *Experimental Wireless (and Wireless Engineer)*, vol. 3, Apr. 1926, p. 203; May 1926, p. 309; July 1926, p. 417; Aug. 1926, p. 483.

4. Studies Relating to the Use of Aluminum Conductors for Pipe-Type Cable, **Joseph Sticher, Louis Meyerhoff, R. H. Hiester, M. H. McGrath.** *AIEE Transactions*, vol. 73, pt. III-B, Oct. 1954, pp. 1126–43.

◆

Discussion

L. E. Sauer (Westinghouse Electric Corporation, Sharon, Pa.): This paper should be of great interest to all users of inductive apparatus as it points up the large amounts of energy that can be stored in inductance coils and sometimes released too quickly during switching operations. Many engineers recognize the need for discharge resistances on generator and motor field switches, but do not see why air-core inductances, current-limiting reactors, should need to be shunted. They seem to think that if there is no iron in the core of the coil there will not be much voltage generated when the circuit is interrupted, forgetting that the real cause of the induced voltage is $d\phi/dt$, the rate of change of flux, and that the iron in a field coil, due to eddy currents, actually slows down the flux change. Even the highly laminated iron circuit of transformers slows down the flux change to approximately five times circuit voltage, whereas a case on record for an air-core inductance, used to limit the starting current of a large motor, flashed over insulation that had a 60-cycle flashover strength of 35 times the circuit voltage. Many coils in present day use as bus-tie current-limiting reactors are above the energy range of the authors' coil, under fault conditions. Consider 46 millihenry charged to 8,000 amperes, 1.45 times the energy; or 34.5 millihenry which may easily be charged to 10,000 amperes, 1.7 times this coil's energy storage.

The authors discussed at length the ratio of a-c to d-c resistance, although the coil proportions and method of winding were selected to give minimum d-c resistance rather than minimum a-c resistance. With reference to the 10-millisecond discharge, their statement that, "...tests of this time duration have shown that the a-c resistance of the coil did not exceed 1 ohm" leads one to ask what was the calculated a-c resistance? Assuming the 0.03 ohm to be at 25 degrees centigrade and the interstrand resistance to be no better than for bare copper, the 162-strand aluminum cable should not exceed a 25-cycle resistance of 0.160 ohm. (A 10-millisecond discharge will effectively approximate 25 cycles.) This assumes individually supported turns; in a cylindrical axial-layer winding of this type, the axial pressure might short-circuit the stranding of as much as the middle third of the winding. However, at 5,200 amperes there appears to be less than 100 pounds of pressure per inch of cable at the central plane, and this is insufficient to cause actual short-circuiting. In dry aluminum cable, i.e., cable not immersed in oil or filled with grease which retards the reforming of oxide rubbed off because of magnetic forces, wound into current-limiting reactors, no increase of eddy loss has been observed due to magnetic forces on the strands. This is observed to hold true up to the point where permanent deformation of the strands takes place, and the short-circuiting mentioned in the foregoing results in an explosion of at least the outside layer of the cable. Enameled-strand-copper cable withstands only about 10% more current density than the bare aluminum before exploding, so for all practical purposes the bare aluminum strands can be considered to have interstrand insulation equivalent to enameled wire. If cable has unidirectional lay of strands, the possibility of explosion is eliminated.

The next thing to consider in looking for extra a-c resistance would be extraneous metals; coil hardware, terminals and connecting bus, building construction, concrete reinforcement, and adjacent apparatus. The metal alone used in supporting the coil-winding form could account for the large a-c d-c ratio.

The use of transformer iron in the core of this coil to produce three times the inductance appears highly optimistic. To do this means that the reluctance of the coil-flux path must be reduced to one third of its present value since the same magnetizing force is to be used. It is proposed to stack transformer iron inside the wooden-coil form only. The inside diameter of the wooden-coil form is not given, but it is probably 6 inches less than the inside of the coil, leaving only 54-inch diameter for "stacking" iron. This 54 inches probably represents area which carries 45% of the coil flux; due to the depth of winding a considerable amount of flux will be within the coil section. If the coil flux has a density of 5,000 gauss in air, this center portion of iron, since the magnetomotive force is the same, will have to carry 245% of the flux in 45% of the area, which gives a density of 27,200 gauss, based on solid iron section. If 5% more flux is added to the air part to cover iron space factor, then 40% iron has 240% flux or 30,000 gauss.

From another viewpoint, the coil dimensions indicate an internal reluctance of only 30% since it is a very short coil. If the internal reluctance is cancelled completely, $1/0.7=1.43$ times the air-core inductance, 40% of the internal area will then have 17,900 gauss, which is attainable but not economically justifiable. Putting the shield on the outside of this coil would be more profitable, but the air density inside the coil would increase the conductor a-c resistance considerably, which in a more frequently used coil might not be allowable.

In considering scale models in the design of inductances, reference 2 of the paper will prove helpful.

Compensation for Instrument Losses in Circuits Containing a Wattmeter, Voltmeter, and Ammeter

NORMAN M. VRANA
ASSOCIATE MEMBER AIEE

THE relative location, in a circuit, of a wattmeter, voltmeter, and ammeter generally requires careful consideration. If the voltmeter and wattmeter potential circuit are connected on the load side of the ammeter and wattmeter current coil, the potential circuits are subjected to the correct voltage, but the ammeter and wattmeter current is in error by the current through the potential circuits. On the other hand, connecting the voltmeter and wattmeter potential coil to the line side of the ammeter and wattmeter current coil permits the correct load current to flow through the instruments, but the potential-circuit voltage is incorrect by the voltage drop in the ammeter and wattmeter current coil.

In both cases the wattmeter reading may be corrected by computing the loss in the potential circuit (or the current circuit, as the case may require) or the losses may be measured by means of a tare reading. Correction for the current-circuit loss varies with the load and is usually less convenient to make. A tare reading of the current-circuit loss requires that the wattmeter potential coil be temporarily connected across the current circuit with the required current flowing. If the correction is to be computed, the resistance of the current circuit is less frequently known than the resistance of the potential circuit so that an additional measurement may be necessary. In those instances when the instrument losses are a significant proportion of the wattmeter reading, an error in the determination of these losses, or in the calibration of the instrument, results in a larger error in the computed load power.

Some of the better quality low power-factor wattmeters contain an internal compensating winding which is connected in series with the potential circuit of the wattmeter. The compensating winding is wound on the current coil so that the effect of the potential-circuit current flowing through the current coil is cancelled by the same current flowing through the compensating winding. This method, however, does not compensate for the voltmeter loss and still permits the ammeter current to be in excess of the load current. Internal wattmeter compensation then leaves something to be desired when the objective is to obtain instrument readings, all of which represent load quantities.

In general, the correction for the ammeter or voltmeter reading, whichever is in error, is laborious though possible. Whenever load phase-angle or a-c resistance is desired, the accuracy may be lowered by errors for which correction is impossible. A means of compensating for these errors so that the instrument readings would all simultaneously represent load quantities is advantageous. In addition, such compensation may permit a lower and more sensitive wattmeter range.

Current-Error Compensation

The circuit shown in Fig. 1 was claimed to compensate for all the errors mentioned. The circuit uses a compensating transformer (an instrument potential transformer or other transformer with low leakage reactance is desirable) connected to the line and arranged to generate a compensating current equal to the total current drawn by the voltmeter and wattmeter potential circuit. The compensating current might be considered to be directed through the ammeter and wattmeter circuit in a direction opposite to that of the actual potential-circuit current.

The resistance R is adjusted with the load circuit open, and its correct value is obtained when the wattmeter reads zero. Under these conditions the same voltage is across both the load terminals and the compensating transformer.

When the load circuit is closed, the current-circuit impedance drop, located in the compensating circuit, increases the magnitude of the compensating current. Furthermore, the current-circuit impedance drop causes the transformer primary voltage to be greater than the load voltage. This latter effect adds to the error in the compensating current.

The error in the wattmeter reading becomes serious for low power-factor loads, either leading or lagging, and may reach a magnitude of 2% or 3%. This is a large enough error in many cases to cause the apparent a-c resistance of an air-cored inductance, computed from the ratio P/I^2, to be less than the d-c resistance.

The magnitude of the error, expressed in watts or in per cent, may be computed from the following relationships developed in the Appendix.

Paper 57-43, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted August 8, 1955; made available for printing November 16, 1956.

NORMAN M. VRANA is with Cornell University, Ithaca, N. Y.

The author acknowledges the assistance received from Robert Z. Fowler who did some of the experimental work on the circuit shown in Fig. 5. At the time, Mr. Fowler was a fifth-year senior student in electrical engineering at Cornell University.

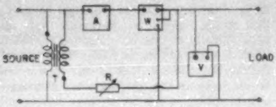


Fig. 1 (left). Circuit which compensates for error in wattmeter and ammeter readings

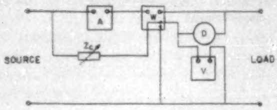
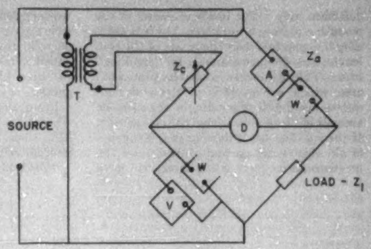


Fig. 2 (left). Circuit which eliminates error in wattmeter and ammeter readings

Fig. 4 (right). Modification of circuit shown in Fig. 3



$$\Delta W = -\frac{(1+a)}{a}|V_1||I_1| \times \left(\frac{R_a \cos\theta_1 + X_a \sin\theta_1}{R_p}\right) \quad (1)$$

$$\frac{\Delta W}{W_1}\% = -\left(\frac{1+a}{a}\right)\left(\frac{R_a + X_a \tan\theta_1}{R_p}\right) \times 100 \quad (2)$$

where

ΔW = watts to be added algebraically to the wattmeter reading to obtain the correct load power
W_1 = load power
$a = N_2/N_1$, the turns ratio of the compensating transformer
R_p = equivalent resistance of the voltmeter and wattmeter potential circuit in parallel
R_a and X_a = resistance and reactance of the current circuit; ammeter and wattmeter current coil in series
V_1, I_1, θ_1 = load voltage, current, and power-factor angle. (Accuracy is sufficient if θ_1 is obtained by using the uncorrected wattmeter reading, but the sign of θ_1 must be retained.)

Bridge-Circuit Compensation

With Potential-Circuit Current

Instead of computing the error in the wattmeter reading, it can be eliminated by means of the circuit shown in Fig. 2. Fig. 3 is the same circuit redrawn in the form of a bridge, which in reality it is. The detector D may be any high-impedance voltmeter of low range such as a vacuum-tube voltmeter. A cathode-ray oscilloscope may also be used for this purpose. Except in the special case when the phase angles of the current circuit and load circuit are equal, the compensating impedance Z_c must contain adjustable resistance and reactance components. Using the criterion for balance in a 4-arm a-c bridge

$$Z_c Z_1 = R_p Z_a$$

$$Z_c = R_p \left|\frac{Z_a}{Z_1}\right| \underline{/\theta_a - \theta_1} \quad (3)$$

Equation 3 indicates that for typically inductive-current circuits, Z_c is capacitive or inductive depending on whether the load phase angle θ_1 is greater than or less than θ_a respectively. For highly inductive loads with circuits operating at 60 cycles, the capacitance required in Z_c for a parallel R-C impedance may be too large for practical purposes. There are at least three modifications of the circuit which can be used to reduce the capacitance required:

1. An air-cored inductance can be inserted in series with the current-circuit arm of the bridge. As long as the Q of the added inductance is greater than the Q of the instrument current coils, the over-all θ_a will be closer to θ_1.

2. An instrument potential transformer can be connected as shown in Fig. 4. The added voltage in the bridge circuit requires that the resistance component of Z_c be large. Therefore, the required capacitance effect can be produced with a relatively small value of parallel capacitance.

3. If the potential transformer shown in Fig. 4 is replaced by a phase shifter, Z_c can be reduced to a pure resistance.

The degree of accuracy available with linear load impedances is a function of the sensitivity of the detector and can be made high. If the load is nonlinear, the harmonics generated cannot be balanced out at the same setting of Z_c for which the fundamental is in balance. In this case, since the detector measures the difference between the load- and potential-circuit voltages, a rough estimate of the worst possible error in the wattmeter reading may be obtained by assuming that the detector harmonic voltage is in phase with its harmonic counterpart in the current circuit.

The bridge circuit is easily balanced with or without the modifications suggested since the reactive and resistive adjustments are located in the same arm. Rebalancing is necessary whenever any of the circuit impedances are altered. The variation of current-circuit impedance with deflection in itself usually has negligible effect on the instrument readings, although the detector may be sensitive enough to indicate the unbalance.

With Line Current

The circuit shown in Fig. 5 uses the line current, a current transformer, and a compensating impedance Z_c to duplicate the voltage drop across the ammeter and wattmeter current coil. The impedance Z_c is approximately equal to the current-circuit impedance times the ratio of secondary to primary turns of the current transformer. An adjustable inductance of low resistance, patterned after the Brooks inductometer,[1] is suitable for use as the reactive component of Z_c. To make the equipment more generally useful and flexible, all coils of the inductometer should be brought out to separate

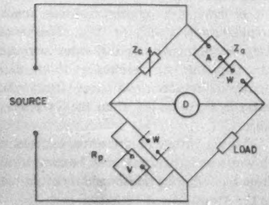


Fig. 3 (left). Same circuit as Fig. 2, redrawn in form of a bridge

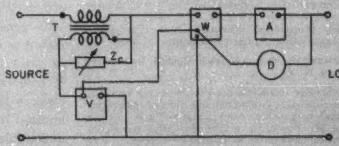


Fig. 5 (right). Circuit which uses line current for compensation

terminals. A decade resistance box with small enough resistance steps or an equivalent adjustable resistance is needed for the resistance adjustment.

Transformer T should be a good quality instrument transformer. A compromise between its turns ratio and the magnitude of Z_c should be made in order to avoid overburdening the transformer. For a 1-to-1 ratio, Z_c is approximately equal to the current-circuit impedance, and the transformer operates within rated burden. For a step-down current ratio of 2 to 1, Z_c is approximately twice the current-circuit impedance, but the transformer may be overburdened enough to introduce a harmonic voltage in the circuit. It has been found that for typical instruments, current ratios lying between a step-up ratio of two and a step-down ratio of two are the limits between too small a value of Z_c for practical construction and the production of harmonic disturbance in the circuit.

Conclusions

1. The circuit shown in Fig. 1 is convenient to use since it relies on only one adjustment of the compensating impedance regardless of load conditions.

2. The author knows of no practical condition in which this circuit would involve an ammeter error which is not negligible.

3. Serious error in the wattmeter may be present in this circuit particularly at low power-factor loads. The error can be calculated by use of equations 1 or 2.

4. If compensation for the wattmeter error is desired rather than calculated corrections, the circuits shown in Figs. 2 and 5 can be used to advantage. The circuits can save time over computational corrections on those occasions when a number of load conditions are to be investigated.

5. The arrangement in Fig. 2 or its suggested modifications will, in general, involve compensating-impedance components which are more common than the compensating impedance used in Fig. 5.

6. The compensating impedance of Fig. 2 varies with the nature of the load impedance, whereas the impedance used in Fig. 5 is a function of the current-circuit impedance. Thus, for the latter circuit an impedance can be constructed for use with a rather large number of different instruments regardless of the nature of the load.

7. Whenever the load has a high degree of nonlinearity, the character of the load is not even approximately defined in simple terms of power factor, resistance, and reactance, and attempts to define the load in such terms will have little meaning. This point is mentioned because the use of a typically sensitive detector in the circuits mentioned will readily show any nonlinearity which otherwise might justifiably be ignored (and undetected) in other arrangements.

Appendix. Derivation of Equations 1 and 2

Referring to Fig. 1, $I_c = \dfrac{a(V_1+I_1Z_a)+I_1Z_a}{aR_p}$

where

V_R = voltage across compensating resistor R
I_p = total potential-circuit current
I_c = compensating current
$I_1 \approx I_1+I_p-I_c$
$I_c = V_R/R$
$R = aR_p$
$V_R = a(V_1+I_1Z_a)+I_1Z_a$

The error current in the current circuit is

$$I_p-I_c=\frac{V_1}{R_p}-\frac{a(V_1+I_1Z_a)+I_1Z_a}{aR_p}$$

$$=-\frac{1+a}{a}\;\frac{I_1Z_a}{R_p}$$

The component of (I_p-I_c) in phase with the potential-circuit voltage V_1, causes the wattmeter reading to be in error. From the preceding equation this component is

$$I_e=-\frac{1+a}{a}\;\frac{|I_1|\;|Z_a|}{R_p}\cos\,(\theta_1+\theta_a)$$

The wattmeter reading will depart from the correct value by the amount V_1I_e. This error in watts is

ΔW = watts to be added algebraically to the wattmeter reading to obtain the correct load power

$$\Delta W=-\frac{1+a}{a}\;\frac{|V_1|\;|I_1|\;|Z_a|}{R_p}\cos\,(\theta_1+\theta_a)$$

The per-cent error is

$$\frac{\Delta W}{W}\times 100=\frac{|V_1|\;I_e}{|V_1|\;|I_1|\cos\theta_1}$$

$$=-\frac{(1+a)}{a}\;\frac{|Z_a|}{R_p}\;\frac{\cos\,(\theta_1+\theta_a)}{\cos\theta_1}\times 100$$

Both of these expressions are made more convenient for computing purposes by expanding the cos $(\theta_1+\theta_a)$ term

$$\Delta W=-\frac{(1+a)}{a}\,|V_1|\;|I_1|\times\left(\frac{R_a\cos\theta_1+X_a\sin\theta_1}{R_p}\right)$$

$$\frac{\Delta W}{W}\%=-\frac{(1+a)}{a}\left(\frac{R_a+X_a\tan\theta_1}{R_p}\right)\times 100$$

Reference

1. A Variable Self and Mutual Inductor, H. B. Brooks, F. C. Weaver. *Paper 290*, U.S. Bureau of Standards, Washington, D.C.

◆

Discussion

R. F. Estoppey (Weston Electrical Instrument Corporation, Newark, N. J.): The method outlined for compensation of instrument losses should certainly be useful when making measurements on circuits of low power. Since the method requires additional equipment and set-up time, the question arises as to when it is necessary to resort to this method.

The approximate power consumed by portable instruments is in the order of 1.0 watt for ammeters and 1.0 watt for the current circuits of the wattmeter. The wattmeter potential-circuit power and voltmeter power are usually less than 3 watts each at 120 volts. If these instruments are used to measure a 120-volt 5-ampere 500-watt load, and the potential circuits are across the source, the voltmeter will be in error by approximately 0.33% and the wattmeter in error by approximately 0.4%. Where accuracy of 0.75% to 1.0% is sufficient, no correction is needed for instrument losses.

Higher voltages and currents usually require the use of potential and current transformers, in which case, the instrument losses become smaller as a percentage of the load, and no correction is necessary. If a low power-factor load such as 100 watts, at 120 volts and 5 amperes, is considered, losses can be quite high. A low power-factor wattmeter can have a loss of 3 watts in the current circuit and 4 watts in the potential circuit. With the potential circuit connected across the source, the voltmeter can be in error by 0.4% and the wattmeter in error by 4%. For potential circuits across the load, the wattmeter is in error by 7% and the ammeter by 0.2%.

These examples point out that it is usually not necessary to make corrections for loads, near unity power factor, of 500 watts or higher. However, at low power-factor loads, instrument losses can become quite high and should be evaluated to determine whether or not compensation or correction is necessary.

No mention has been made of a-c instrument amplifiers which have been developed to supply the instrument losses. The potential amplifier is made to have a very high input impedance, and the current amplifier has a very low input impedance. By using one or both amplifiers, the instrument losses can be neglected. These amplifiers can be quite costly and are usually used where manual correction for instrument losses is not possible or desirable.

N. M. Vrana: The techniques suggested in this paper are directed primarily toward making it possible to obtain results which make significant use of the guaranteed accuracy (0.25% to 0.5%) of portable laboratory instruments. Without computed corrections or compensation of some kind, this objective is impossible of attainment particularly when the power factor is low, as Mr. Estoppey has stated. The need for utilizing instrument accuracy becomes especially important when the instrument readings are to be used to compute impedance, a-c resistance, and power factor, since the computational process increases the degree of uncertainty in the results.

Whenever it is desired to increase the accuracy of the results of any measurement, the time required to make the measurement increases and usually the equipment becomes more expensive. The techniques suggested are no exception to this fact, but the circuit components required are simple to work with and are typically available or readily made for subsequent use as discussed in the paper.

Transistor 2-Terminal Switches

ABRAHAM HAR'EL
NONMEMBER AIEE

Synopsis: A 2-terminal self-biased memory-type switching device is an electronic bistable switch which has only two accessible terminals to all voltage and current sources. Such a switch can be realized with a single multiterminal semiconductor device exhibiting negative resistance, if advantage is taken of the "blocking voltage" concept. The blocking voltage is defined here as the forward voltage at a P-N junction at which the incremental conductance is a hundredfold greater than the incremental conductance at zero voltage. By proper circuit connections, the blocking voltage may be utilized to block injection of minority carriers in a semiconductor device until the voltage applied to the device exceeds a peak value. Such a device may exhibit low or negative a-c resistance after the peak value is exceeded. It is shown in the following discussion how the blocking-voltage concept may be applied to obtain negative resistance behavior from a junction transistor, a point-contact transistor, a modified double-base diode, or a modified tetrode. A few applications, mainly in the telephone field, are also indicated.

IN A centralized telephone system, all telephones are connected to and disconnected from each other through a central office. The interconnection between subscribers' lines is performed in what is called a "switching matrix," shown schematically in Fig. 1. This matrix enables any line to be connected on demand to any other line through a common link; no more than two lines at a time can be connected to a link. The number of links used in the matrix depends upon the expected volume of traffic in the office. The electric connection at the intersections between lines and links is through an electronic switch called a "crosspoint." Crosspoints must meet many switching and transmission requirements including the following:

1. Because there is generally a large number of crosspoints in any office, the size, number of components, and d-c power consumption per crosspoint should be minimized.

2. Each crosspoint should be a memory-type switch, i.e., a bistable element, in order to simplify and economize the control circuits.

3. The crosspoint must present a very high d-c and a-c impedance in its OFF state, and a very low a-c impedance in its ON state.

4. As a bidirectional element in the transmission path, the crosspoint must not introduce appreciable linear or nonlinear distortion or noise. It may possibly serve as an amplifier.

5. It should be reliable, stable, and inexpensive.

The simplest approach to crosspoint design is to use a 2-terminal element. The main advantage of a 2-terminal element is that its behavior depends only on the voltage difference between its two terminals and not on the potential of each terminal with respect to a fixed reference point. An element with three or more terminals generally does not have this property. Because of certain impedances in the line and link circuits, the d-c potentials on the line and link change as a result of turning a crosspoint on, thus making the crosspoint "float" above ground potential. This would change the d-c bias conditions on the third terminal. Therefore, the crosspoint should be self-biased, i.e., it should not include a d-c bias source.

The five requirements can be met by a semiconductor crosspoint, that need consist only of a single semiconductor device and a few passive elements, if the crosspoint has a voltage-current characteristic with negative-resistance region between the ON and OFF states.

Certain gas tubes have the desired characteristic (Fig. 2). However, the research program at the Moore School, sponsored by the U. S. Signal Corps, is specifically interested in semiconductor devices, which, when used in a proper circuit, exhibit a similar characteristic.[1]

The purpose of this paper is to describe a circuit which inherently meets the requirements specified in the foregoing, and to point out its applications. The discussion will include operation with a single-junction transistor, a single point-contact transistor, or a single modified double-base diode.

Blocking-Voltage Concept

The circuits discussed in the following sections are all based on a blocking-voltage concept which is associated with all semiconductor P-N junctions.

It is well known[2] that the current through a P-N junction is governed by the exponential law

$$I=I_s(e^{qV/kT}-1)$$

The numerical value of the bracketed term, at room temperature, is shown as a function of voltage in the following table.

|V| Volt	+ |V| $(e^{V/0.026}-1)$	− |V| $(e^{-V/0.026}-1)$
0	0	0
0.02	1.16	−0.536
0.04	3.65	−0.785
0.06	9.0	−0.900
0.08	21	−0.955
0.1	46	−0.979
.	.	.
.	.	.
.	.	.
0.2	2,200	−1.0
0.3	100,000	−1.0

The saturation current I_s is very small, a few microamperes in germanium diodes, and a small fraction of a microampere (μa) in silicon diodes. The diode current in the reverse direction reaches its saturation value at 0.1 volt; see section on the junction transistor switch. The diode current in the forward direction is given by the product of the saturation current and the exponential factor. It is seen from the table that, at forward voltages which do not exceed 0.1 volt, the diode current is less than 100 times the saturation current. As the saturation current is very minute, of the order of 1 μa or less, the forward diode current at these voltages is very small, of the order of 100 μa or less, i.e., practically cut off. The forward voltage at a P-N junction, at which the incremental conductance is a hundredfold greater than the incremental conductance, at zero voltage is defined here as the blocking voltage. Thus, the blocking voltage is of the order of 0.1 volt at germanium junctions and 0.5 to 1.0 volt at silicon junctions, on the assumption that 100 μa or less is cut-off current in the circuits discussed in the following.

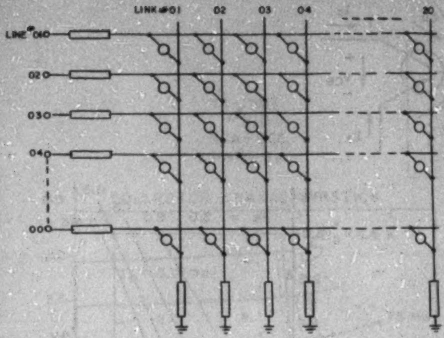
Instead of regarding this voltage as an undesired voltage drop across a conducting diode, it is planned to utilize it as a useful voltage, which, by proper circuit connections, can block injection of minority carriers in a diode even when biased in the forward direction. It is shown in the following sections that a voltage many times greater than the blocking voltage

---

Paper 57-106, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted September 19, 1956; made available for printing February 21, 1957.

ABRAHAM HAR'EL is with the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa.

This work is part of a research project sponsored by the U. S. Signal Corps. The author wishes to thank the sponsor and the following colleagues: C. Eldert, L. N. Kanal, A. J. Messey, T. R. Williams, H. Yamada, and particularly L. M. Spandorfer and M. H. Korzen, for their contributions and co-operation.



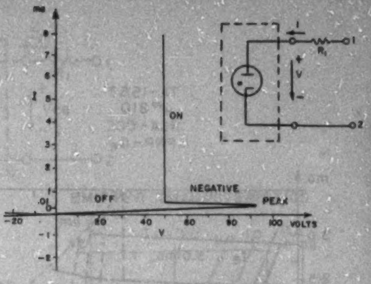


Fig. 1 (left). Switching matrix

Fig. 2 (above). Gas-tube circuit and characteristic

can be applied to the 2-terminal circuit in the forward direction, and can establish a forward voltage across the active diode which is smaller than the blocking voltage, thus keeping the current through the diode smaller than 100 μa, i.e., effectively cut off. The foregoing table shows that, while the current is very small for voltages smaller than the blocking voltage, the current increases very rapidly at slightly higher voltages. The higher the blocking voltage, the higher the voltage that can be applied to the circuit in the forward direction and still keep it cut off. Therefore, these circuits will naturally operate better with silicon devices than with germanium devices, provided the negative-resistance properties are comparable.

Point-contact diodes exhibit a similar behavior. The saturation current, however, is greater, hence the blocking voltage is lower.

Junction-Transistor Switch

The blocking-voltage concept can be applied to a junction transistor in the design of an interesting circuit. Recently, a few papers have been published which throw a new light on transistor design and transistor applications.[2–5] These papers discuss the use of a junction transistor as a circuit element exhibiting useful negative resistance, and they refer to earlier theoretical studies[6] which predicted the existence of avalanche multiplication in semiconductors and in P-N junctions. Since February 1955, development work along similar lines has been under way at the Moore School, particularly in the development of the 2-terminal self-biased circuit utilizing the negative resistance of junction transistors and other devices.[1]

Before describing in detail the operation of the switching circuit based on the negative resistance of the junction transistor, it is worth while to review the characteristics of a P-N-P junction transistor. Fig. 3 shows, at the left-hand side, the collector current versus collector-to-base voltage. The solid lines represent constant emitter currents. Usually the family of curves would not be shown beyond 20 volts for this transistor. Collector current increases gradually with collector voltage and, at 20 volts, collector currents equal emitter currents. This is the locus of the points of zero base current, and is shown at the $I_b=0$ curve. As collector voltage is increased still further, collector current exceeds emitter current. The operating points do not exceed the 75-milliwatt maximum permissible power dissipation, shown by the dotted line. The voltages are kept far below the reverse breakdown voltage of the base-to-collector diode.

In the transistor under consideration, and in the voltage range shown, the increase of collector current with voltage is rather small, and collector current exceeds emitter current only slightly. A small positive current flows into the base; the magnitude of this current is the amount by which collector current exceeds emitter current. The base currents range from zero to a few hundred microamperes. Fig. 3 shows the locus of the points where collector current exceeds emitter current by 0, 20, 40, and 60 μa. These constant positive base-current characteristics are plotted with dashed lines. Observe the negative slope of these characteristics. The negative slope exists because the absolute increase of collector current with increased collector voltage grows rapidly with increased current.

On the right-hand side of Fig. 3, the family of emitter characteristics is shown where the parameter is collector current. Notice the curve of zero collector current. This is essentially a forward characteristic of a P-N junction diode: the emitter-to-base diode (open collector). The curve shows clearly the blocking voltage, which, in this case, is about 0.1 volt; at lower voltages the current is very small, and at higher voltages it rises very rapidly. The constant collector-current curves usually do not extend into the region where emitter current is smaller than collector current; but in this transistor they do, by an amount which corresponds to the positive base currents. With the help of these two families of characteristics, it is possible now to proceed to the graphical analysis of the circuit which uses this transistor.

Fig. 4 shows how these curves are used to determine the voltage-current characteristics of the circuit which is shown at the top of this figure. This is the basic switching circuit discussed in this and the following sections. The resistor R_b, which connects base to emitter, is of prime importance and largely determines the operating characteristics of the circuit. Suppose that the resistor R_b is 6.8 kilohms. This resistor determines the emitter-to-base voltage for any base-to-collector current. The voltage in turn determines the forward emitter current according to the emitter characteristics so that the sum of the emitter, collector, and base currents is always zero. This in turn defines the operating points in the emitter and collector characteristics. By continuing this procedure, point by point, one gets the operating curve, which is shown by the heavy line, in the emitter family and in the collector family. As an example, three typical operating points are computed for $R_b=6.8$ kilohms.

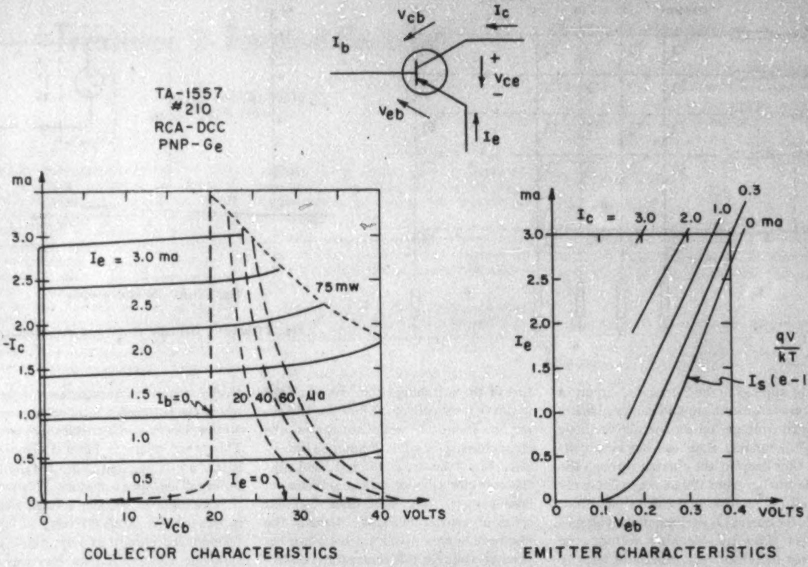


Fig. 3. Junction-transistor characteristics

Operating point *A*
$I_b=10\ \mu a$
$V_{eb}=I_b.R_b=0.068$ volt
I_e is very small (cut off)
I_c is very small (cut off)

Operating point *B*
$I_b=18\ \mu a$
$V_{eb}=I_b.R_b=0.120$ volt
$-I_c=I_e+0.018$ ma (milliampere)
$V_{eb}=0.120$ volt
$-I_c \doteq 0.1$ ma
$-V_{cb} \doteq 34$ volt

Operating point *C*
$I_b=27\ \mu a$
$V_{eb}=I_b.R_b=0.184$ volt
$-I_c=I_e+0.027$ ma
$V_{eb}=0.184$ volt
$-I_c \doteq 2$ ma
$-V_{cb} \doteq 24$ volt

These operating points *A*, *B*, *C* are typical operating points in the OFF and ON high-negative-resistance regions respectively. Thus, a typical switching characteristic is obtained which includes a useful negative-resistance state.

It is important to realize that the resistor R_b keeps the base current virtually constant at high emitter currents even though the resistor is relatively small. The voltage of the emitter characteristic is constant and therefore determines a constant base current. Constant base current is the most convenient way of obtaining useful negative resistance from the junction transistor.

The experimental voltage-current characteristic of this circuit agrees very closely with this graphical method; both yield the operating curve shown by the heavy line. This switching circuit, using a single junction transistor, and its measured operating characteristic are shown in Fig. 5.

In its OFF state, the circuit presents 1- to 10-megohm impedance both to alternating and direct current. In its ON state it consumes about 50-milliwatt d-c power and presents a negative resistance of about 1 kilohm at audio-frequencies; the reactance at these frequencies as well as the nonlinear distortion is small. Thus, the transmission properties of the circuit are satisfactory, and the circuit can serve as a bistable flip-flop and as a gated transmission path, simultaneously.

It should be emphasized that there is no bias source in the emitter-to-base circuit. The d-c bias and the switching pulses are applied to terminals 1 and 2 only. The transmitted a-c signal is also applied to the same two terminals which present very high or low a-c impedance to the signal, according to the state in which the device resides. Suppose the device is in the OFF state. An audio signal applied to terminals 1 and 2 will not develop any current in load R_1 unless the signal is preceded by a starting pulse of the proper polarity and sufficient amplitude which turns the switch ON. Any signal applied afterward passes freely through the closed switch and develops the desired a-c power in the load. When a terminating pulse of opposite polarity and sufficient amplitude is applied, the switch is turned OFF again, blocking further audio transmission. Thus a 2-terminal self-biased memory-type transmission and switching element has been achieved.

Before concluding the discussion of the junction transistor, a few remarks should be made in connection with the avalanche effect and how it affects the peak points of the constant base-current characteristics and the operating characteristic of the 2-terminal circuit.

Fig. 6 shows the collector characteristics of another junction transistor. The graph at the top of the figure shows collector current versus collector-to-base voltage, for constant positive base currents. It has already been mentioned that the negative slope of these curves results from the fact that the absolute increase of

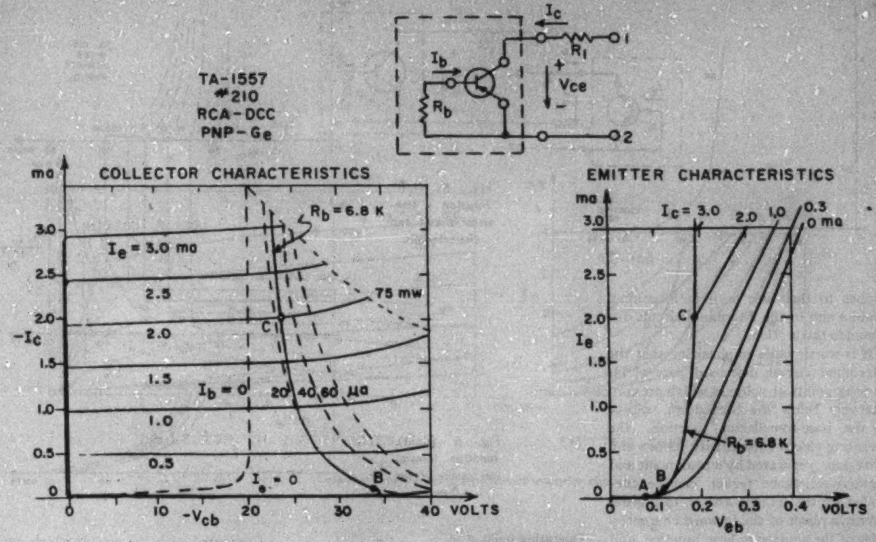


Fig. 4. Graphical analysis of junction-transistor circuit

collector current with increased collector voltage grows rapidly with increased emitter current. There are a number of factors which may cause this increase of collector current. The most important ones are the body-avalanche effect, nonlinear surface conduction, impure junctions, and Zener breakdown. In various transistors, one or another effect predominates and consequently determines the shape of the characteristics.

When the body-avalanche effect is the predominant factor, it is shown in the previously mentioned references that the collector current is subject to a multiplication factor M, which is approximately given by the relation

$$M=\frac{1}{1-\left(\frac{V}{V_B}\right)^n}$$

where V is collector-to-base voltage, V_B is the avalanche breakdown voltage, and n depends on the type of semiconductor. For holes in germanium, n is approximately three.

The last statement can be reversed and one may say that when it is found that the collector current is subject to the multiplication factor M, the predominant factor in that transistor is a pure body-avalanche effect. However, in many cases, collector current increases according to a different law which implies effects besides the body-avalanche effect.

This is of particular importance when the collector characteristic is examined for open emitter, which means that the collector-to-base diode is examined in its reverse direction. The center diagram in Fig. 6 shows, in solid line, the measured voltage-current characteristic of the collector-to-base diode in its reverse direction. The dot-dashed line in the same diagram represents, for the same scale, the theoretical voltage-current characteristic of the same collector-to-base diode had it followed the pure body-avalanche law

$$\frac{I_0}{1-\left(\frac{V}{V_B}\right)^3}$$

where $I_0=1$ μa and $V_B=50$ volts. The actual characteristic, shown in solid line, indicates clearly that in the voltage range of 20 to 45 volts this transistor is affected by factors other than the pure body-avalanche effect. Probably, the predominant factor in this case is a nonlinear surface conduction. There is an obvious correlation between this base-to-collector characteristic and the peak points of the constant base-current curves of the same transistor, as shown in the top and center diagrams. For example, the peak point of the $I_b=60$ μa curve is 33 volts in the top diagram, and it corresponds to the point of 33 volts and 60 μa on the characteristic shown in the center diagram. Had the characteristic been a pure bulk-avalanche curve, as shown by the dot-dashed line in the center diagram, the constant base-current curves would have resulted in the shape shown by the dot-dashed lines in the bottom diagram. All peak points would have been concentrated near the break-point, at voltage V_B. Only voltages higher than 97% of the break voltage V_B can yield base currents which are greater than ten times the base-to-collector saturation current, if the transistor is subject to a pure avalanche effect.

There are cases where it is desired to have a fixed single peak point in a transistor, which is independent of the magnitude of base current, and which yields a maximum voltage range for the negative-resistance region. In such cases one would use a pure body-avalanche transistor. On the other hand, where it is desired to have variable peak points in a transistor, which can be controlled by the magnitudes of base currents, one must use a transistor which shows the effect, demonstrated in Fig. 6. This discrimination between the two types of transistors

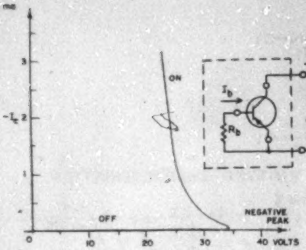


Fig. 5 (left). Junction - transistor circuit and characteristic

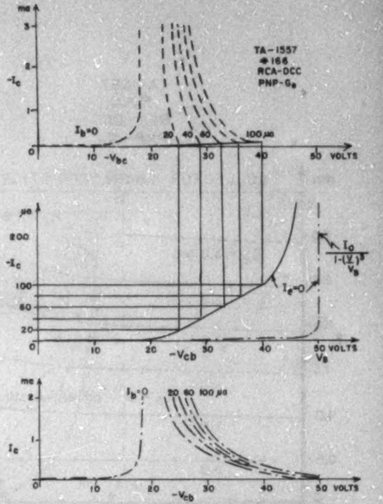


Fig. 6 (right). Junction transistor: peaks versus $I_e=0$ curve

applies to their use both in 3-terminal circuits and in the 2-terminal circuit discussed in this section.

It is worth while emphasizing that the 2-terminal circuit discussed here yields its peak points at voltages which are considerably below the breakdown voltage of the base-to-collector junction; the operating characteristic, with its OFF and ON regions, separated by a peak point and negative-resistance region, can be controlled by the resistor and is, to a great extent, a result of the forward characteristics of the emitter-to-base junction, and mainly its blocking voltage.

Point-Contact Transistor Switch

The blocking-voltage concept can be similarly applied in the design of a negative-resistance 2-terminal unit using a point-contact transistor.[7,8] The transistor connection and related graphs of the original and derived operating characteristics are shown in Fig. 7. Both the collector and emitter characteristics of a point-contact transistor are shown. The collector characteristics are usually plotted for constant emitter currents, shown in solid lines. In this figure, however, the curves of constant base currents are also shown, in dashed lines, superimposed on the constant emitter-current curves, for the same scale. Notice the curve $I_b=0$. This is the locus of the points where $I_c=I_e$, and is nearly a vertical line, with constant collector voltage of about -0.5 volt. The curves of constant base currents have a negative slope. Those shown in the figure are 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 ma. They are plotted according to

$$I_b+I_c+I_e=0$$

With the help of these curves and the emitter characteristics, it is possible to plot the operating points for any R_b. Suppose that $R_b=100$ ohms, including the effective internal base resistance. Two typical operating points are computed here as an example.

Operating point *A*
$I_b=1.0$ ma
$V_{eb}=I_b . R_b=0.1$ volt
$-I_c=I_e+1.0$ ma
$V_{eb}=0.1$ volt
$-I_c \doteq 1.2$ ma
$I_e \doteq 0.2$ ma
$-V_{cb} \doteq 6.8$ volts

Operating point *B*
$I_b=2.5$ ma
$V_{eb}=I_b . R_b=0.25$ volt
$-I_c=I_e+2.5$ ma
$V_{eb}=0.25$ volt
$-I_c \doteq 4.2$ ma
$-I_e \doteq 1.7$ ma
$-V_{cb} \doteq 4.2$ volts

The operating points *A* and *B* are typical points in the OFF and ON (negative-resistance) regions, respectively. Similarly, the entire curve representing the voltage-current characteristic of this circuit can be plotted. Two such curves are shown, one for $R_b=100$ ohms and one for $R_b=200$ ohms. The higher the resistance, the lower the voltages. Fig. 8 shows the switching circuit with a typical operating characteristic.

It is assumed here, and in the preceding section, that V_{cb} can be substituted for V_{ce}. This is justified because of the relatively low emitter-to-base voltages. Another assumption is that resistance R_b stays constant. However, in the case of the point-contact transistor, that part of R_b which represents the internal base resistance is appreciable and varies with high-current densities. This explains the slight differences between the experimental and theoretical curves.

Compared with the circuit which employs a junction transistor, this point-contact transistor circuit uses a smaller base resistor R_b, the peak points occur at lower voltages, the sustaining voltages in the ON region are very low, and the ratio of peak to sustaining voltages is much higher; the OFF impedance is much smaller, and the ON a-c impedance is also smaller, turning positive sometimes.

Double-Base Diode and Tetrode Switch

A similar approach can be applied also to a double-base diode[9] and a tetrode.[10] The double-base diode is a 3-terminal semiconductor device with a single P-N junction and is capable of exhibiting a useful negative-resistance region.

The conventional circuit of the double-base diode and its voltage-current characteristic are shown in Fig. 9. The emitter characteristic has three regions: OFF, NEGATIVE, and ON. In the OFF region, emitter current is cut off even though the emitter is a few volts more positive than base 1. The base 2 to base 1 voltage establishes in the N bar, opposite the P junction, a potential which is more positive than the emitter potential; therefore, the junction is biased in the reverse direction. Keeping V_b constant and increasing V_e, the P-N junction is finally biased in the forward direction; minority carriers are freely injected into the N bar and diffuse towards base 1. The voltage re-

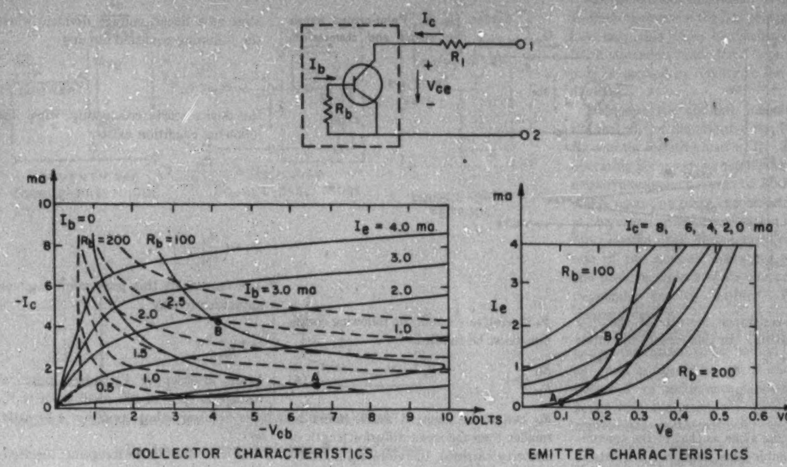


Fig. 7. Graphical analysis of point-contact transistor circuit

quired to sustain this ON current is very low and the resistivity of the N bar drops. Between the peak and the ON region there is a negative-resistance region.

So far, 3-terminal operation of the conventional double-base diode has been described. However, one is tempted to suggest a slight modification in the construction of the double-base diode so that it may be operated as a 2-terminal self-biased memory-type switch.

Fig. 10 shows the modified version. Base 2 is placed very close to the junction and is connected directly (or through external impedance Z) to the emitter lead. Both bias and signal are applied to terminals 1 and 2. It was mentioned previously that, when the forward bias on a P-N junction is less than the blocking voltage, only a minute current flows in the forward direction. If base 2 is very close to the junction, the positive d-c voltage applied between terminals 1 and 2 can reach 10 volts or so before blocking voltage is exceeded at the junction. An OFF region is thereby present. At higher voltage, a large number of minority carriers are injected from the emitter toward base 1, and the sustaining voltage becomes very small; the negative resistance and ON regions are thereby present.

This qualitative description of the operation of the modified double-base diode can be accompanied by a mathematical presentation. The notation used is that shown in Fig. 10. The junction forward characteristic V_j-I_j, instead of being completely idealized (zero voltage drop), is assumed to have a constant voltage drop equal to the blocking voltage V_o; $I_j=I_e$, $V_j=V_b-V_e$. Analysis of the regular double-base diode operation[11,12] shows that the negative-resistance region of its voltage-current characteristic is given approximately by the equations:

$$V_e=-\frac{\mu_n}{\mu_p}\cdot R_{b1}\cdot I_e+R_{b1}\cdot I_b \qquad \text{, at } I_b=\text{constant} \tag{1}$$

$$V_e=-\frac{\mu_n}{\mu_p}\cdot\frac{R_{b1}\cdot R_{b2}}{R_{b1}+R_{b2}}\cdot I_e+\frac{R_{b1}}{R_{b1}+R_{b2}}\cdot V_b, \qquad \text{at } V_b=\text{constant} \tag{2}$$

where μ_n, μ_p are electron and hole mobilities respectively, and R_{b1}, R_{b2} are base 1 and base 2 resistances (to emitter) respectively.

The special condition for the modified double-base diode is

$$V_e=V_b-V_o$$

where V_o is the blocking voltage which may be regarded as a small internal bias battery between emitter and base 2, preventing the emitter from injecting any appreciable amount of minority carriers until its external potential V_b exceeds that of the adjacent base region V_e by more than the blocking voltage V_o, that acts as a contact potential at the interface between the P and N regions of the junction. By substituting this condition in equation 2,

$$V_b-V_o=-\frac{\mu_n}{\mu_p}\cdot\frac{R_{b1}\cdot R_{b2}}{R_{b1}+R_{b2}}\cdot I_e+\frac{R_{b1}}{R_{b1}+R_{b2}}\cdot V_b$$

the following results

$$V_b=-\frac{\frac{\mu_n}{\mu_p}\cdot\frac{R_{b1}\cdot R_{b2}}{R_{b1}+R_{b2}}}{1-\frac{R_{b1}}{R_{b1}+R_{b2}}}\cdot I_e+\frac{1}{\frac{R_{b1}}{R_{b1}+R_{b2}}}\cdot V_o$$

This can be simplified into

$$V_b=-\frac{\mu_n}{\mu_p}\cdot R_{b1}\cdot I_e+\left(\frac{R_{b1}}{R_{b2}}+1\right)\cdot V_o$$

By changing the notation to that used in Fig. 10

$$V_b=V$$

$$I_e=I-I_b$$

the equation assumes the form

$$V=-\frac{\mu_n}{\mu_p}\cdot R_{b1}\cdot(I-I_b)+\left(\frac{R_{b1}}{R_{b2}}+1\right)\cdot V_o$$

However, the following condition keeps the base current virtually constant, similar to the base current in the transistor 2-terminal circuit; see the section "Junction Transistor Switches" and Fig 4.

$$I_b=\frac{V_o}{R_{b2}}$$

Hence

$$V=-\frac{\mu_n}{\mu_p}\cdot R_{b1}\cdot I+\left[\frac{R_{b1}}{R_{b2}}\left(\frac{\mu_n}{\mu_p}+1\right)+1\right]\cdot V_o \tag{3}$$

This final equation shows clearly the

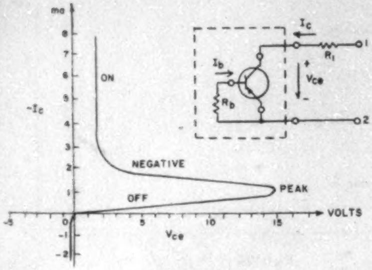

ma
8
7
6
5
4
3
2
1
-1
-2
$-I_c$
ON
NEGATIVE
PEAK
OFF
VOLTS
-5
0
5
10
15
V_{ce}
I_c
I_b
R_1
R_b
+
V_{ce}
-
1
2

Fig. 8. Point-contact transistor circuit and characteristic

negative-resistance feature of the *V-I* characteristic. In this region, the input resistance is

$$r_{in}=\frac{\partial V}{\partial I}=-\frac{\mu_n}{\mu_p}\cdot R_{b1}$$

which is the same as that of the conventional double-base diode with constant base-current bias.

In order for the device to be realizable and have cut-off negative-resistance and saturation regions, it is necessary that within a finite range, i.e., a few volts and a few milliamperes, equation 3 yields operating points with positive voltage ($V>0$) and positive current ($I>0$). As V_o is relatively small, the following condition must be fulfilled

$$\frac{R_{b1}}{R_{b2}}>>1 \tag{4}$$

R_{b1} cannot be changed much (must be smaller than the mean diffusion length of minority carriers), therefore R_{b2} must be made as small as possible.

Designate as V_p that value of V which must be applied in order to develop the blocking voltage V_o across the P-N junction. This is the peak voltage at which the device leaves its cut-off region and enters its negative-resistance region. Then, regarding the base in the cut-off state as a linear voltage divider, where the following equation obtains

$$V_j=\frac{R_{b2}}{R_{b1}+R_{b2}}\cdot V$$

the device starts conducting when the following condition exists

$$V_o=\frac{R_{b2}}{R_{b1}+R_{b2}}\cdot V_p$$

i.e.,

$$V_p=\left(\frac{R_{b1}}{R_{b2}}+1\right)\cdot V_o$$

The current at this peak point is given approximately by

$$I_p=\frac{V_p}{R_{b1}+R_{b2}}$$

Again, in order for V_p to be a practical voltage in spite of the small value of V_o, it is necessary that condition 4 be satisfied.

As was stated in the foregoing, the negative resistance cannot extend beyond the region where $I>0$ and $V>0$, from which it follows that

$$\left[\frac{R_{b1}}{R_{b2}}\left(\frac{\mu_n}{\mu_p}+1\right)+1\right]\cdot V_o>\frac{\mu_n}{\mu_p}\cdot R_{b1}\cdot I$$

Hence, neglecting the 1 in the brackets, the negative resistance cannot extend

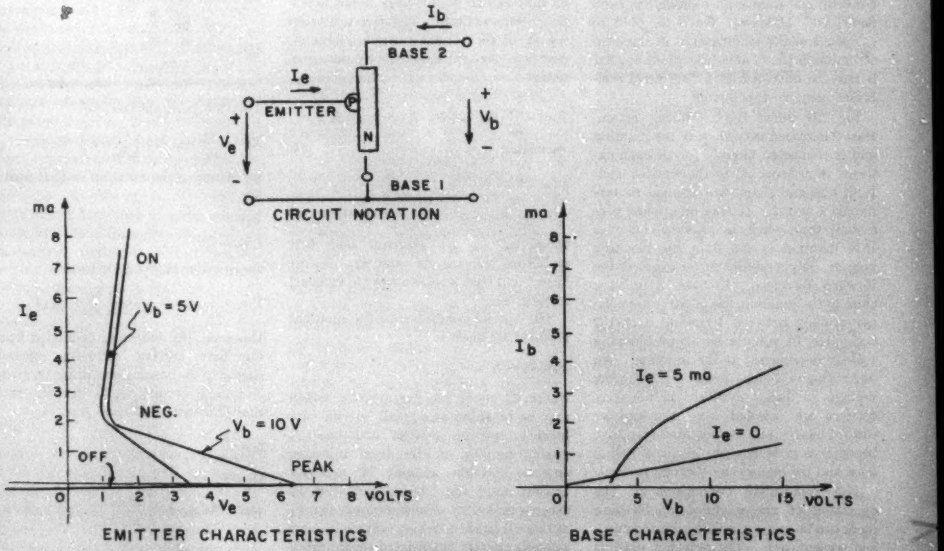
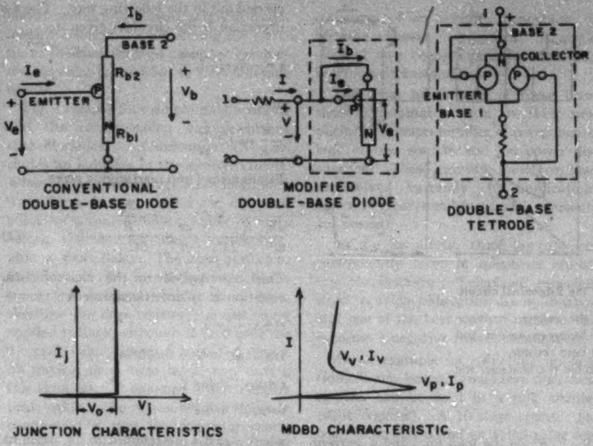

I_b
BASE 2
I_e
EMITTER
P
N
+
V_e
-
+
V_b
-
BASE 1
CIRCUIT NOTATION
ma
8
7
6
5
4
3
2
1
I_e
ON
V_b = 5V
NEG.
V_b = 10 V
OFF
PEAK
0
1
2
3
4
5
6
7
8 VOLTS
V_e
EMITTER CHARACTERISTICS
ma
8
7
6
5
4
3
2
1
I_b
I_e = 5 ma
I_e = 0
0
5
V_b
10
15 VOLTS
BASE CHARACTERISTICS

Fig. 9. Double-base diode characteristics



I_b
BASE 2
I_e
R_b2
+
V_b
−
+ EMITTER
P
N
V_e
R_b1
−
CONVENTIONAL
DOUBLE-BASE DIODE
I_b
I
I_e
1o
+
V
−
P
N
V_e
2o
MODIFIED
DOUBLE-BASE DIODE
1
+
BASE 2
N
COLLECTOR
EMITTER
P
P
BASE 1
2
DOUBLE-BASE
TETRODE
I_j
V_o
V_j
JUNCTION CHARACTERISTICS
I
V_v, I_v
V_p, I_p
MDBD CHARACTERISTIC

Fig. 10. Modified double-base diode and tetrode

beyond the valley-point current I_v, which is given by

$$\left(1+\frac{\mu_p}{\mu_n}\right)\cdot\frac{V_o}{R_{b2}}>I_v$$

This applies, of course, only to a straight-line approximation which is based on extrapolation of the negative-resistance formula beyond the boundaries in which it is valid. In this straight-line approximation the negative-resistance region extends from $(I_p,\ V=V_p)$ to $(I_v,\ V=0)$.

The following is a numerical illustration. If

$V_o=0.05$ volt
$R_{b1}=10$ kilohms
$R_{b2}=50$ ohms
$\dfrac{\mu_n}{\mu_p}=2$

then

$V_p=10$ volts
$r_{in}=-20$ kilohms, in the negative-resistance region
$r_{in}'=10$ kilohms, in the OFF region
r_{in}'' is very low, in the ON region
$I_p=1.0$ ma
$I_v=1.5$ ma

Negative-Resistance Device as Circuit Element

Each of the circuits described in the foregoing is a 2-terminal circuit whose d-c voltage-current characteristic contains a region of negative incremental resistance (negative a-c resistance). The general behavior of 2-terminal negative-resistance devices as circuit elements is discussed rather extensively in the literature.[13]

As shown in Fig. 11, there are two basic groups of negative-resistance devices: the *S* type (open-circuit stable, current controlled) where voltage is a single-valued function of current, but current may be a multivalued function of voltage; and the *N* type (short-circuit stable, voltage controlled) where current is a single-valued function of voltage, but voltage may be a multivalued function of current.

All the devices mentioned in the previous sections are of the *S* type, and therefore only this type is discussed here. Each statement made for *S*-type devices may be applied by duality to the *N*-type devices.

Designate the absolute magnitude of the real part of the negative resistance by R_o, and the resistive load by R. A stable operating point can be established in the negative-resistance region of $R>R_o$. No stable operating point can be established in the negative-resistance region if $R<R_o$ but, instead, the device may reside in one of two possible operating points in the positive regions of its characteristic, depending on its "history" (bistable memory device).

An *S*-type negative resistance always causes the current to lag behind the voltage, and its operation can be understood better if an inductance is assumed in series with the resistance. This can be represented mathematically as follows: The negative-resistance part of the characteristic may be written as

$$V=E_o-R_o\cdot I$$

and if $V=V(t)$, $I=I(t)$, then

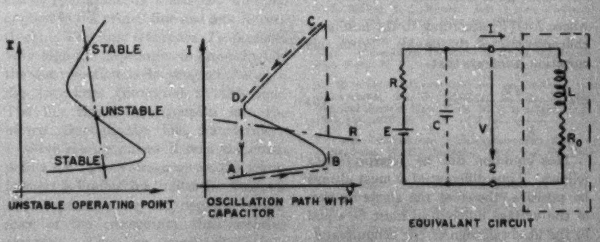
$$V(t)=E_o-R_o\cdot I(t-t_o)$$

Expanding by Taylor's series gives

$$I(t-t_o)=I(t)-t_o\frac{dI(t)}{dt}+\frac{t_o^2}{2!}\frac{d^2I(t)}{dt^2}-\dots$$

Neglecting higher order terms and substituting gives

$$V(t)=E_o-R_o\cdot I(t)+R_o\cdot t_o\cdot\frac{dI(t)}{dt}$$



I
"N" TYPE
V
I
"S" TYPE
V
I
STABLE
R
STABLE OPERATING POINT V
I
STABLE
UNSTABLE
STABLE
UNSTABLE OPERATING POINT
I
C
D
R
A
B
OSCILLATION PATH WITH CAPACITOR
I
1
R
E
C
V
L
R_o
2
EQUIVALANT CIRCUIT

Fig. 11. Negative resistance

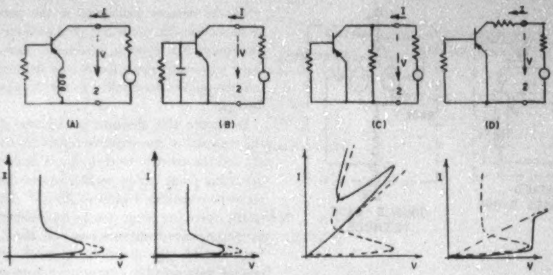


Fig. 12. Modifications of the 2-terminal circuit

A—Includes an inductor in the emitter circuit
B—Capacitor is connected between emitter and base, shunting the base resistor
C—Different type, relates to the d-c characteristic
D—Same as C

Designate

$$R_o \cdot i_o = L$$

Thus

$$V(t) = E_o - R_o \cdot I(t) + L\frac{dI(t)}{dt} \qquad (5)$$

The equivalent circuit, which includes an inductance in series with the negative ohmic resistance, is shown in Fig. 11. This inductance is important, not only for stability considerations, but mainly from a point of view of transmission. While in the negative-resistance region, the device may be used to pass an a-c signal through it; on one hand this a-c signal is amplified, and on the other hand it is affected by this equivalent inductance which is proportional to the magnitude of the negative resistance.

If a capacitor C is connected across the terminals of the negative resistance, the circuit may oscillate, and the path of oscillations in the v-i plane is shown in Fig. 11. This path may be determined from the following equation

$$E = V + C\cdot\frac{dV}{dt} + IR$$

where $I = f(V)$ according to the nonlinear characteristic of the device. From this equation it follows that

$$dt = C\frac{dV}{\dfrac{E-V}{R} - f(V)} > 0 \qquad (6)$$

This equation can be interpreted as follows. Time differential dt must always be positive, therefore the algebraic sign of the voltage differential must be equal to the algebraic sign of the denominator. For any voltage V there exists a current $E - V/R$, on the load line, and a current $I = f(V)$, on the characteristic. Thus, for any abscissa, if the load-line ordinate is greater than the characteristic ordinate, their difference is positive and therefore dV is positive, causing the operating point to move to the right, from A to B. If the load-line ordinate is smaller than the characteristic ordinate, their difference is negative and therefore dV is negative, causing the operating point to move to the left, from C to D. The same consideration holds for the discontinuity points.

A general discussion of an S-type negative resistance with a capacitor C across it, connected through a load R to a voltage source E, which analyzes the stability of the circuit, as well as its transient behavior and type of a-c oscillations, may be carried out in the following way. The a-c operation of the circuit is given by

$$v = -R_o\cdot i + L\frac{di}{dt}$$

and

$$-\frac{v}{R} = i + C\frac{dv}{dt}$$

Eliminating i and rearranging gives

$$\frac{d^2v}{dt^2} + \frac{\dfrac{L}{R} - R_oC}{LC}\cdot\frac{dv}{dt} + \frac{1 - \dfrac{R_o}{R}}{LC}\cdot v = 0 \qquad (7)$$

This corresponds to the characteristic equation (Laplace transformed)

$$s^2 + as + b = 0$$

Its solution is of the form

$$i = a_{11}\epsilon^{s_1t} + a_{12}\epsilon^{s_2t}$$

$$v = a_{21}\epsilon^{s_1t} + a_{22}\epsilon^{s_2t}$$

where v and i are referred to an origin at the point of equilibrium, i.e., the d-c portion of the solution is omitted. The a_{ij} coefficients depend on both the initial conditions and the circuit parameters. Only when the real part of both s_1 and s_2 is negative, the circuit is stable. There are six possible solutions for the response of this circuit to a disturbance: decreasing exponential or oscillation (stable), increasing exponential or oscillation (unstable), constant-amplitude sine-wave oscillations and constant-amplitude multivibrations.

Modifications of Original Circuit

The preceding semiconductor circuit which can be operated as a 2-terminal self-

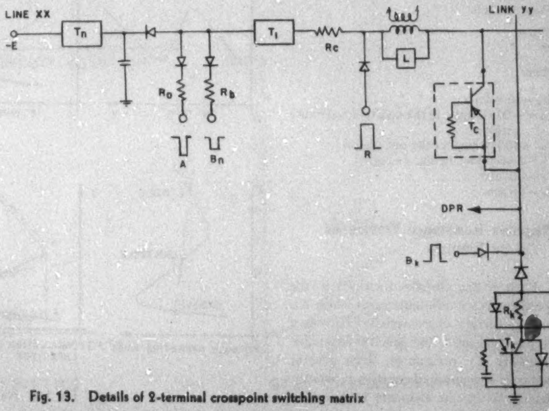


Fig. 13. Details of 2-terminal crosspoint switching matrix

biased memory-type switching and transmission element, by connecting the emitter to the base through an ohmic conductor, may also be modified to yield different results.

Fig. 12 shows four simple modifications and the corresponding voltage-current characteristics. Modification (A) includes an inductor in the emitter circuit. When the applied voltage has a fast-rising wave form, it may exceed the d-c peak point by a considerable number of volts before the emitter starts conducting, after a short delay. The same applies to modification (B), in which a capacitor is connected between emitter and base, shunting the base resistor. A fast-rising applied voltage, although it surpasses the d-c peak value, does not turn the device ON until a short time later, and during this time delay it assumed a much higher peak value. The reason for this delay in conduction is that the capacitor short-circuits out the voltage developed across the resistor, thus preventing it from exceeding the blocking voltage for a short time.

Modification (C) is of a different type and relates to the d-c characteristic. When a resistor is connected across the 2-terminal element, the composite characteristic assumes the shape shown in the figure. It is plotted by adding the ordinates (currents) of the two individual characteristics at a common abscissa (voltage), because the two elements are in parallel. The resulting characteristic is not only multivalued in current, for a single-valued voltage, but is also multivalued in voltage for a single-valued current. The latter mode may be regarded as the dual of the first mode, and thus the same single device may be operated in two dual modes. The characteristic contains two negative-resistance regions and three positive-resistance regions.

Modification (D) is of the same kind as modification (C). The resistor is connected in series with the 2-terminal device. The composite characteristic, which can be seen in the figure, is plotted by adding the abscissas (voltages) for a common ordinate (current) of the two individual characteristics. It is possible to eliminate the negative resistance entirely this way, and have a nonlinear positive resistance composed of three distinct slopes.

Telephone Switching Matrix and Other Applications

There is a wide variety of applications for these transistor 2-terminal switches. They may be used wherever a negative-resistance of an astable, monostable or bistable element is required. A single element may be used both for switching and for a-c transmission.

The great advantages of these transistor 2-terminal circuits are their simplicity, minimum number of components, lack of bias source, and low-power dissipation. These circuits, therefore, lend themselves naturally to applications where a great number of such elements are needed.

As a-c amplifiers, these circuits can conveniently serve in telephone repeaters, for instance. Telephone repeaters must provide bidirectional amplification and one of the best ways to achieve this is to use a negative-resistance amplifier.[14]

As an example for switching applications, it may be mentioned here that these circuits can be used in a ring counter (shift register). A 10-stage counter has been built and tested as a dial-pulse register in a telephone-switching system.[1] The advantage of this register is that it needs a minimum number of components per stage; it uses the junction transistor 2-terminal circuit.

Another example of the use of these circuits shows how they are used both as switches and as transmission elements simultaneously. As mentioned previously, such are the requirements for crosspoints in a telephone switching and transmission matrix,[1] shown in Fig. 1. The details of such a crosspoint using the junction-transistor 2-terminal circuit, with the associated link and link circuits, are shown in Fig. 13. Ordinarily, the crosspoints T_c are in their OFF state. When a line calls, a negative pulse A is applied which turns on one, and only one crosspoint on this line, connecting it to an idle link. (This is assured by the competing-lockout operation, and the lockout resistor R_a.) The dial pulses develop a voltage across the link resistor R_k, and are routed into the dial-pulse register *DPR*. The crosspoint between the called line and the seized link is turned on by simultaneous pulses B_n and B_k, applied to the proper line and link respectively. The link transistor T_k presents very high a-c impedance to ground while the conversation is in progress between the two lines connected to this link. The line signals are coupled into the matrix through the line transformers. Positive release pulses R turn the crosspoints off. The voltage across the transformers is limited by limiter L. Resistor R_c compensates for the negative resistance of the crosspoints, thus assuring stability. The link transistor T_k is a junction transistor biased to 6 ma, providing the proper impedance to ground. T_n is the line junction transistor biased to 3.3 ma, providing the proper load line. The other junction transistor T_1 serves as a current limiter and is biased to 5 ma.

Summary

The foregoing discussion has explained the blocking-voltage concept and has shown how it can be utilized to design transistor 2-terminal switches. The emphasis has been on the forward characteristics of the semiconductor diode. It has been shown that when a resistor shunts the diode, a controlling voltage is developed. This voltage maintains the device in its high-impedance state as long as it is smaller than the blocking voltage. After an applied pulse switches the device into its high-conduction state, the controlling voltage cannot increase appreciably above the blocking voltage, and constant current bias is thus maintained.

This method can be used to design a 2-terminal self-biased memory-type switching and transmission element, which is composed of a single semiconductor device combined with one ohmic resistor. Circuits have been discussed where the single semiconductor device is a junction transistor, a point-contact transistor, a modified double-base diode, or a modified tetrode. Each of these devices must exhibit a negative-resistance region. Several applications have been discussed, in particular a telephone switching matrix.

References

1. Dial Telephone Switching Circuits. *Quarterly Progress Reports*, Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa., 1955, 1956.

2. The Theory of P-N Junctions in Semiconductors and P-N Junction Transistors, W. Shockley. *Bell System Technical Journal*, New York, N. Y., vol. 28, July 1949, pp. 435–89.

3. Delayed Collector Conduction, A New Effect in Junction Transistors, M. C. Kidd, W. Hasenberg, W. H. Webster. *RCA Review*. Princeton, N. J., vol. 16, Mar. 1955, pp. 16–33.

4. Alloyed Junction Avalanche Transistors, S. L. Miller, J. J. Ebers. *Bell System Technical Journal*, vol. 34, Sept. 1955, pp. 883–902.

5. Junction Transistor Electronics, J. L. Moll. *Proceedings*, Institute of Radio Engineers New York, N. Y., vol. 43, Dec. 1955, pp. 1807–19.

6. Electron Multiplication in Silicon and Germanium, K. G. McKay, K. B. McAfee. *Physical Review*, New York, N. Y., vol. 91, Sept. 1953, pp. 1079–84.

7 The Transistor, a Semiconductor Triode, J. Bardeen, W. H. Brattain. *Ibid.*, vol. 74, July 1948, pp. 230–31.

8. Physical Principles Involved in Transistor Action, J. Bardeen, W. H. Brattain. *Ibid.*, vol. 75, Apr. 1949, pp. 1208–25.

9. The Double-Base Diode: A New Semiconductor Device, I. A. Lesk, V. P. Mathis. *Convention Record*, Institute of Radio Engineers, Mar. 1953, pp. 2–8.

10. A Junction Transistor Tetrode for High Frequency Use, **R. L. Wallace, Jr., L. G. Schimpf, E. Dickten.** *Proceedings*, Institute of Radio Engineers, vol. 40, Nov. 1952, pp. 1395–1400.

11. Double Base Expands Diode Applications, **J. J. Suran.** *Electronics*, New York, N. Y., vol. 28, Mar. 1955, pp. 198–202.

12. Low-Frequency Circuit Theory of the Double-Base Diode, **J. J. Suran.** *Transactions*, Professional Group on Electron Devices, Institute of Radio Engineers, vol. ED-2, no. 2, Apr. 1955.

13. Dynamics of Transistor Negative-Resistance Circuits, **B. G. Farley.** *Proceedings*, Institute of Radio Engineers, vol. 40, Nov. 1952, pp. 1497–1508. (Additional references in this paper.)

14. Negative Impedance Telephone Repeaters, **J. L. Merrill, Jr., A. F. Rose, J. O. Smethurst.** *Bell System Technical Journal*, vol. 33, Sept. 1954, pp. 1055–92. (Additional references in this paper.)

Magnetic Circuits with D-C Magnetization and High Alternating Induction

H. L. GARBARINO
ASSOCIATE MEMBER AIEE

Synopsis: A fairly simple theory is given and criteria have been established for analysis or design of magnetic circuits in which there are an unbalanced or d-c magnetization and a high alternating induction. High values of alternating induction and a nonlinear relation between induction **B** and field strength **H** are characteristics of magnetic circuits in power apparatus, in contrast to magnetic circuits in audio components. The relation between induction and field strength during cyclic variation is given approximately by the d-c magnetization curve of the magnetic circuit. Maximum and minimum values of induction correspond to maximum and minimum values of field strength, but the average values of **B** and **H** are not simply related. The demagnetizing effect of an alternating field superimposed upon a direct field is the reduction of average induction after the application of the alternating field. This effect is due to the nonlinearity of the magnetization function. A simple circuit using the test sample and an auxiliary transformer has been devised for obtaining data on magnetic cores by direct measurement. Test results for representative cores have been organized in a manner suitable for design use. Excitation volt-amperes, optimum per-cent gap, and core loss may be found as a function of a-c induction and unbalanced field strength. The data on magnetic materials are applicable to reactors and to transformers with unbalanced direct current such as units for supplying half-wave rectifiers. Analysis of the frequency components aids in understanding the operation of half-wave supply circuit. Only the induced voltage of the primary winding and the average secondary current affect the magnetic circuit.

THE PURPOSE of this paper is to present criteria for the design of magnetic circuits in which there are a combined d-c magnetization and a high alternating induction. Electromagnetic devices in which there is an unbalanced magnetic force present difficulties as a result of the nonlinear relationship between induction or flux density and magnetic field strength. This factor is of less importance when either the unbalanced magnetization or the superimposed alternating induction is small, because existing design methods for such cases are reliable and widely used, principally in the design of inductance units and audiotransformers for electronic circuits. In such cases inductance is almost independent of the magnitude of the alternating induction. In power apparatus, however, alternating induction is always much larger, and the nonlinear characteristics of the material make it impractical to apply algebraic methods for obtaining relations among the magnetic quantities. The quantities are the instantaneous values of induction or flux density and the magnetizing force. Other important variables are the configuration of the magnetic circuit, the type of joint used, and the grade and thickness of the material.

The superposition of d-c and alternating components of magnetizing force on magnetic materials gives characteristics which are qualitatively similar in some respects to the case of an alternating magnetization alone. In both instances the magnetic quantities may be related by a hysteresis curve, which is usually given with induction or **B** as ordinate and magnetizing force or **H** as abscissa. With an unbalanced magnetization, the curve is dissymmetrical. In addition, the average values of **B** and **H** are not simply related to each other by the d-c magnetization characteristic of the material. In power apparatus, the hysteresis loop is relatively narrow, with or without an unbalanced magnetization, so that the relation between **B** and **H** is practically single-valued. Therefore the maximum and minimum values of induction and magnetizing force respectively are approximately related by the d-c magnetization characteristic.

The parameters which must be determined in the design of magnetic circuits are: suitable material, value of flux density, proportions for the magnetic circuits, and the nonmagnetic gap in the core. It is desirable that the resulting device be as small as possible for the required power rating. This means that operating densities—flux density in the core and current densities in the windings—should be made as high as design limitations will permit. Such limitations may be core or winding losses, voltage regulation, efficiency of the unit, or permissible heating. The most universal limitation in power and communication equipment is permissible temperature rise. Regardless of other important specifications to be met, excessive temperature must be avoided. A characteristic of almost all power equipment is that temperature rise increases as size decreases, for a given output. To obtain maximum rating with least material, the designer should approach the permissible operating temperatures of both core and winding. The quantities which determine whether the magnetic circuit has been properly designed are the core loss and the magnetizing reactive power required by the core. Either of these quantities may limit flux density in the core. In physically small apparatus, the contribution of the magnetizing power to the heating of the windings is usually the important design limitation, while in large equipment, core loss is usually the limitation.

The analysis and data to follow are quantitatively applicable to magnetic circuits in which there is an unbalanced

---

Paper 57-599, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Great Lakes District Meeting, Des Moines, Iowa, April 15–17, 1957. Manuscript submitted October 16, 1956; made available for printing March 20, 1957.

H. L. Garbarino is with the Armour Research Foundation, Illinois Institute of Technology, Chicago, Ill.

This paper is based upon part of an investigation conducted for the Signal Corps Engineering Laboratories and Wright Air Development Center. The support of these organizations and the assistance of their technical staff members are gratefully acknowledged. This material is part of a thesis submitted to Illinois Institute of Technology.

magnetization and a sinusoidal variation in induction or flux density. This implies a sinusoidal power source and a negligible resistance in the winding which excites the magnetic core. Applications are transformers and reactors which carry an unbalanced direct current. Such an unbalance may be a recognized feature of a design, such as in a transformer supplying a half-wave rectifier, or it may represent a malfunctioning of some associated equipment, such as the failure of one rectifier in a full-wave output circuit.

A device very closely related to the given examples is the saturable reactor or magnetic amplifier. Qualitatively, performance characteristics are similar to the transformer or reactor with unbalanced magnetizing force. The basic difference is that the saturable reactor is employed with relatively large impedance in series with the winding which provides an alternating excitation to the magnetic core.

Background References

There have been many publications on unbalanced magnetization of magnetic circuits, which generally fall into the following categories: basic characteristics of magnetic materials, methods of magnetic measurement, design of transformers and reactors which carry unbalanced direct current, saturable reactor circuits or magnetic amplifiers, and analyses of nonlinear circuits. Two extensive bibliographies have been published by Rex[1] and by Miles,[2] both primarily intended to present a background on magnetic amplifiers.

Many investigators have found that losses in magnetic cores are increased by an unbalanced magnetization, among them Vallauri,[3] Rosenbaum,[4] Holm,[5] Chubb and Spooner,[6] and Ball.[7] Schunck[8] studies the effect of an unbalance on wave shapes of flux, current, and voltage, and the reduction in inductance caused by the unbalance. He also showed that average induction and average magnetizing force are not simply related by the magnetization curve, a fact which is still not always appreciated. Important contributions were made by Niwa and others,[9,10] who found that for a certain d-c magnetization the superposition of an alternating field may either raise or lower the average value of induction. At low field strengths, an alternating flux increases apparent permeability, while at high field strengths, an alternating flux reduces apparent permeability. Another study of the effects of a superposed alternating field on permeability and losses is that of Spooner.[11] Tests on silicon steels have been made by Sims and Clay[12] and results are given for incremental permeability, current wave shapes, and losses at relatively high values of alternating induction and d-c magnetization. More recent tests have been made by Nolen and Niesten,[13] Banerjee,[14] Battelle Memorial Institute,[15] and Armour Research Foundation.[16]

Circuits for obtaining losses and effective inductance are in two general classes: null-balance or a-c bridge types, and direct measurement types. In the bridge circuits, a coil on the magnetic sample constitutes one arm of the bridge. The d-c magnetizing force can be supplied through the same winding or by a second winding on the sample. Landon[17] has used the Hay bridge, which is characterized by a series capacitance and resistance in the arm opposite from the inductance, and resistances in the other two arms. This method is compared with the Maxwell and Anderson bridges. Harris and Williams[18] discuss an improved Maxwell bridge. Salmon[19] analyzes the Owen and Anderson bridges. Wilhelm[20] has used a Maxwell bridge with an a-c and d-c supply impressed across the bridge. Other references on bridges are Rehfisch and Bissmire,[21] Welsby,[22] Harris,[23] Dieterly and Ward,[24] and Ryder.[25] Charlton and Jackson[26] have presented a circuit for direct measurement, using two similar cores with two windings on each. Teuchert[27] used a comparative method, whereby a resistance standard is varied to draw the same current as that drawn by the inductance to be measured. Sims[28] proposed the use of one core with two windings, each excited with both alternating and direct current. A circuit given by Nowotny and Söchting[29] consists of one core and two windings, one with a-c and one with d-c excitation, a method which requires a high impedance in the d-c circuit. Daniels[30] describes a method for obtaining a high impedance with an electronic circuit.

Many references deal with the characteristics and the design of transformers and chokes where one winding carries direct current. In 1909, Coales[31] devised a circuit to provide an inductive load for testing alternators. Two windings of two cores were connected in parallel for use as the load and two additional windings were connected in series to a d-c control source in such a way that the induced a-c potentials cancel. This circuit appears to be the first application of a d-c controlled inductance used as a load element. Coales found that inductance decreases and loss increases with the magnitude of the direct current. In 1927, Hanna[32] published a widely used article on design of reactances and transformers, which relates inductance, direct current, magnetic field, core geometry, and air gap in the core. Curves presented by Hanna make it possible to select an optimum air gap. Most of the data and design methods published subsequently are applicable only to much smaller a-c densities than would be used in power apparatus. Magnetic and material data and analyses have also been given by Howe,[33] Minton and Maloff,[34] Turner,[35,36] Gürtler,[37,38] Lanchester,[39] Beatty,[40] Cosens,[41] Baggally,[42] Partridge,[43] Glazier,[44] and Lee.[45,46] Legg[47] has extended the method of Hanna by devising an orderly design method for transformers and reactors carrying unbalanced direct current. Further analysis of this problem is given by Carter and Richards,[48] who also show that average induction and average magnetic field strength are not simply related unless a-c induction of density is small. Inductance design charts which account for winding resistances and turns are given by Crowhurst.[49] An important factor in the design of transformers supplying a half-wave rectifier is the type of output filter circuit. Schade[50] has given widely used curves which relate transformer secondary and load quantities according to frequency, load resistance, filter capacitance, and circuit resistance. Seely[51] has studied the interesting half-wave rectifier output circuit consisting of an inductance and resistance in series.

Many of the reference works on saturable reactors of magnetic amplifiers provide qualitative information applicable to magnetic circuits having an unbalanced magnetization and an alternating induction. One of the early applications for saturable reactors was for the measurement of direct currents, making use of the change in a-c impedance of a winding when another winding on the core is carrying direct current. Circuits are described by Ryan,[52] Besag,[53] Krämer,[54] and Specht and Wagner.[55] The theory and use of d-c controlled reactors for various applications is described by Boyajian.[56] Johnson and Merrell[57] show how induction and magnetic field strength are related for a sinusoidally varying induction, and give design data. They also point out that a superimposed alternating induction tends to demagnetize the core, in that the average induction is thereby reduced.

Smith[18] compares many basic magnetic amplifier circuits and gives algebraic relations among alternating induction, average induction, and average field strength.

General Relations

The nonlinear characteristics of ferromagnetic materials cause considerable complication in an analysis of magnetic circuits with unbalanced magnetization. The problem consists essentially of relating induction or flux density **B** and magnetic field strength **H**. A typical d-c magnetization curve is given in Fig. 1. When the material is subjected to alternating values of **B** and **H** about some average values, the normal curve describes the approximate performance. Qualitatively, there are two rather different conditions: comparatively small variations and comparatively large variations. References which give extensive algebraic relations among the circuit quantities cover the first case for small variations in **B** and **H**.

In the second case, where the variation in induction **B** is large, the d-c curve describes fairly well the relation between **B** and **H**, a fact which is used for the following qualitative analysis. During operation under unbalanced conditions, the exact function consists of a displaced (nonsymmetrical) hysteresis loop, enclosing an area proportional to losses per cycle, as for magnetic materials operated without an unbalance. With a large variation in **B**, the quantity corresponding to incremental permeability is some average slope of the d-c characteristic, and this magnitude is greatly affected by the magnitude of the variation in density **B**. This dependence makes the second case more complicated than the first, inasmuch as incremental permeability is no longer approximately constant, as for low values or alternating density. Therefore the term incremental permeability has little significance for the second case. It is found that the most useful means for understanding the problem are: a graphical analysis of the magnetic quantities and a study of frequency components of the electrical quantities.

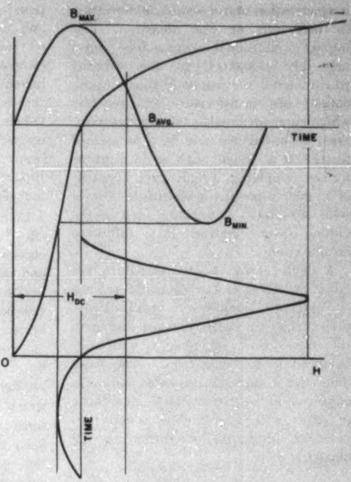
The important variables which determine the characteristics of a magnetic circuit are the average and variable components of induction **B** and magnetic field strength **H**, the geometry and material of the magnetic circuit, and the frequency of variation. Average flux density is defined as $\mathbf{B}_0$, and the average field strength is defined as $\mathbf{H}_{DC}$. A sinusoidal component of flux density is assumed with the maximum value **B**. This is one half of the maximum variation about the average $\mathbf{B}_0$. Important geometric parameters of the magnetic circuit are the dimensions and the type of joint used. It is assumed that the flux density is uniform in the core at every instant of time. In general, the presence of a joint in the magnetic circuit makes the field strength nonuniform around the magnetic circuit. Therefore the given average field $\mathbf{H}_{DC}$ is an average around the core circuit as well as being an average in time. This quantity can be related directly to the direct currents in the windings surrounding the core. If there is a direct current I_{DC} in one winding of N turns then the average field strength may be defined as

$$\mathbf{H}_{DC} = \frac{0.4\pi N I_{DC}}{m} \text{ oersteds} \qquad (1)$$

where m is the mean length of magnetic circuit in centimeters.

Average flux density is not a simple function of the average field strength $\mathbf{H}_{DC}$. The relation between these quantities is given by Fig. 1. If an average value of flux density $\mathbf{B}_0$ is assumed, with a superimposed variation **B**, then the resulting function of field strength **H** is uniquely determined by the d-c magnetization curve, in so far as the hysteresis effect can be neglected. The average of the **H** function must be $\mathbf{H}_{DC}$. This establishes a relation between $\mathbf{B}_0$ and $\mathbf{H}_{DC}$. Average flux density cannot be readily determined from the magnetic characteristic by graphical means, since repeated trials would be required, assuming each time a certain value of average density. In a closed magnetic circuit it is also difficult to determine average density by experimental methods. The only possible way is to trace the magnetic history of an initially demagnetized specimen.

Fig. 1. Relations between induction and field strength



From the foregoing discussion it is found that the magnetic variables of a given magnetic circuit are uniquely determined by the average and varying components of flux density. Since both of these together determine the unbalanced magnetization $\mathbf{H}_{DC}$, it can be reasoned that the performance could also be described by the values of **B** and $\mathbf{H}_{DC}$, which are readily determined.

Study of Fig. 1 shows that the actual value of $\mathbf{B}_0$ is smaller than it would be if **B** were zero. This is an illustration of the demagnetizing effect which has been noted by several investigators. Smith[18] shows this algebraically, and it can readily be confirmed from Fig. 1, that for a constant value of $\mathbf{H}_{DC}$, $\mathbf{B}_0$ decreases when **B** is increased, and conversely. Also, if $\mathbf{B}_0$ were held constant, then $\mathbf{H}_{DC}$ increases when **B** is increased. A demagnetizing effect is obtained only when the slope of the d-c curve decreases for an increasing **H**, or when the second derivative of the **B-H** function is negative. If the slope were increasing instead, the presence of an alternating induction component **B** would increase the value of $\mathbf{B}_0$. However,

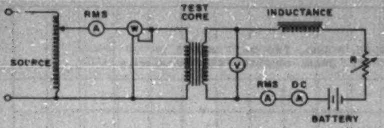


Fig. 2 (left). Circuit for magnetic tests requiring a large inductance

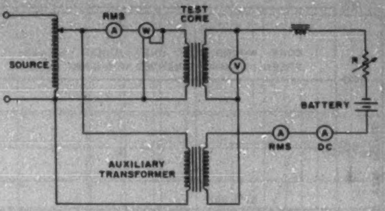


Fig. 3 (right). Circuit for magnetic tests using an auxiliary transformer

there is no advantage in finding the magnitude B_0, since it is only indirectly related to the more readily measurable quantities. Similarly there is no need to find the values of the maximum and minimum induction.

Effect of a Nonmagnetic Gap

In addition to the other physical properties of a closed magnetic circuit, core geometry and material, an important variable is the nonmagnetic gap which may be introduced in some cases to obtain improved characteristics. A core joint of any type has an effect on the core properties which is similar to that of the nonmagnetic gap. In most applications of magnetic cores to power apparatus it is desired that the self-inductance of a given winding on the core have the highest possible value. When there is no unbalanced magnetization present, the highest self-inductance is obtained when a core joint of minimum reluctance is used. If the d-c magnetization curve of Fig. 1 is considered to be a plot of flux density or induction **B** against average magnetic field strength **H** around the entire magnetic path then the introduction of a nonmagnetic gap increases the value of **H** for each value of **B**, or bends the curve to the right.

A gap may yield an increase in self-inductance when a core has an unbalanced magnetizing current in one winding, which can be demonstrated qualitatively. The change in the magnetization characteristic of the core then tends to reduce the average flux density $\mathbf{B_0}$ of the core. The quantities **B** and $\mathbf{H}_{DC}$ are considered constant and independent. The maximum flux density $(\mathbf{B}+\mathbf{B_0})$ is decreased and so is the maximum field strength corresponding to $(\mathbf{B}+\mathbf{B_0})$. The cost of obtaining this advantage is a decrease of slope of the **B**-**H** curve in the regions of low **H**. In another sense, instantaneous permeability (actually a variable during a cycle) is increased at high values of **H** and decreased at low values of **H**. Since it is apparent that a sufficiently large gap would decrease self inductance in any case, there may be some optimum value of nonmagnetic gap, which depends on **B**, $\mathbf{H}_{DC}$, core material and geometry. In a transformer with unbalanced magnetization, the requirement of maximum self-inductance is equivalent to a requirement for minimum magnetizing current in the winding which provides a-c excitation to the core.

The quantities **B** and $\mathbf{H}_{DC}$ are used to predict magnetic characteristics of various cores of various shapes and sizes. Similarly, it is desirable to express nonmagnetic gap in a manner such that the results for one size and shape can be used for others. It can be readily shown that per-cent nonmagnetic gap should be used, or the ratio of air-gap length to magnetic-circuit length. Although the ratio of gap-length to magnetic-circuit length is most important, another consideration is the minimum effective gap length obtainable in particular cores. This effective gap length is present because of the necessity for joints, and depends very little on the size of the core. Typical effective gap lengths for various joints range from 0.0005 to 0.003 inch.

Because of difficulties associated with bridge circuits, a method of direct measurement is preferred for obtaining basic data on magnetic cores and materials, even though it may not yield the sensitivity obtainable from a null-balance circuit. A suitable degree of accuracy in measurements on magnetic materials is achieved when the measurement error is a small fraction of the variability of the material among identical samples. From this standpoint a direct method should give reasonable absolute and relative accuracies for typical data. A direct measurement can be made on a sample which has either one or two windings. When one winding is used, a-c and d-c excitation sources are interconnected. Adjustment of the input and maintenance of a sinusoidal a-c wave shape can be obtained only with testing apparatus of unusual versatility. Therefore a sample with two windings should be considered. The simplest example of a circuit with two windings is shown in Fig. 2. Ideally, the inductance in the d-c circuit should be infinite, such that no alternating current flows in the d-c winding. When this situation is obtained, the total core excitation at any time is the sum of the instantaneous primary current and the d-c secondary current. The power measured by the wattmeter is composed of the primary winding losses and the core losses, while the battery in the secondary circuit supplies the resistive losses of the secondary winding, the choke, and the control resistance. Practically, it is desirable to use sufficient impedance such that power transferred from the a-c winding is negligible. This could be accomplished with a choke many times the size of the sample under test, and suitable for carrying the direct current. An alternative is to measure the a-c real and reactive power components in the d-c circuit.

A solution to the problems presented by the circuit of Fig. 2 is provided by the circuit shown in Fig. 3. An alternating voltage can be introduced in the d-c circuit in such a manner as to oppose the voltage induced in the d-c winding around the test sample. This is accomplished with an auxiliary transformer which need have only the same turns ratio as that of the sample, and be capable of carrying the direct current. With this circuit there will still be a relatively small net alternating voltage in the d-c circuit which is due to distortion of the induced voltages from a sinusoidal wave shape. This in turn is caused by resistances in the primary circuits of both test sample and auxiliary transformer. The net induced voltage will be of harmonic frequencies. However, since the net voltage is of relatively low magnitude and of higher frequencies, a small inductance in the d-c circuit is adequate to make the a-c secondary current negligible.

The circuit of Fig. 3 is actually the same as that described by Coales,[21] which was used as an inductive load for alternator tests. It has apparently not been used to obtain data on magnetic materials. The circuit is also the same

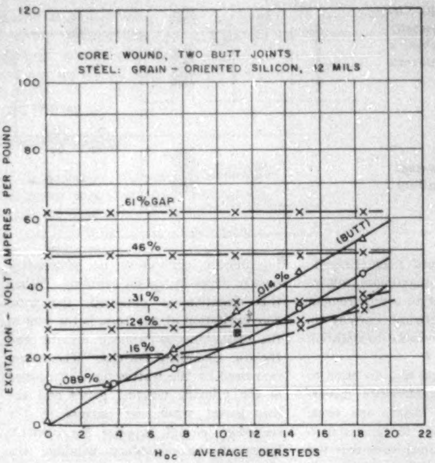


Fig. 4. Excitation of wound core at 80 kilolines per square inch (60 cps)

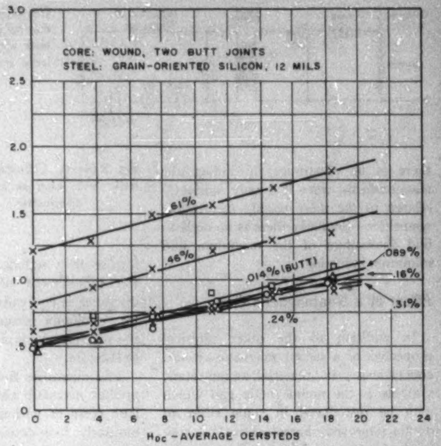


Fig. 5. Core loss of wound core at 80 kilolines per square inch (60 cps)

as that of a simple, parallel-connected amplifier in which load resistance is very small and control circuit impedance is high. The major difference between the magnetic amplifier and the test circuit is that the flux density variation in the magnetic amplifier is not sinusoidal. Also, the maximum current corresponding to maximum **H** in the magnetic amplifier is determined approximately by instantaneous applied voltage and load resistance, rather than by maximum **B**, as in the test circuit of Fig. 3.

Test Results and Design Curves

The variables which have been studied experimentally to compile data for design use are a-c or incremental flux density **B**, average magnetic field strength $\mathbf{H}_{DC}$, length of nonmagnetic gap, grade of core material and thickness, core geometry, and frequency of the power supply, representative of parameters used in small electronic power transformers. In using the circuit of Fig. 3, it would be desirable to measure core loss directly with the wattmeter. However, the window areas of typical small cores are not sufficiently large to accommodate the wire sizes of the a-c winding that would make winding losses negligible. This is a disadvantage which could be overcome by using a much smaller ratio of core cross section to window area. It is less likely that this difficulty would be encountered on much larger cores, because flux density is then limited by core loss rather than by exciting current. With the typical small cores selected, it is easy to measure winding resistance and to calculate winding loss. This is subtracted from total input losses to obtain core loss. It is desirable to use a low-power-factor-type wattmeter for magnetic measurements.

Descriptions of two of the cores tested and graphical results are given in the Appendix. These data are considered to be typical, but since only one sample of each was tested, the variations which would exist among cores of the same type and construction are not known. Figs. 4 and 5 give excitation (volt-amperes per pound) and core loss (watts per pound) for the one core at a typical a-c flux density. The abscissa $\mathbf{H}_{DC}$ is defined by equation 1. The parameter is the effective per-cent nonmagnetic gap, or ratio of effective gap to length of the magnetic circuit.

Results of the tests have been studied to find if the gap which gives minimum excitation also yields minimum core loss. If the values of nonmagnetic gap for minimum core loss and minimum excitation are appreciably different, then both should be considered in selecting the nonmagnetic gap. In general, there is a fair correlation of conditions for minimum core loss and minimum excitation current. That is, minimum core loss and excitation, at the various values of density **B** and magnetization $\mathbf{H}_{DC}$, are usually obtained with the same, or almost the same effective gap. There is better correlation between the conditions for minimum total loss and minimum excitation current. Total input loss in these tests is the sum of core loss and primary winding loss. Since core proportions and winding sizes in the experimental transformers are typical of those that might be used in production units, it is indicated that designing for minimum excitation will tend to yield minimum total losses, even though core losses are not quite at the minimum in all cases.

In accordance with the foregoing discussion, it is desirable to select a nonmagnetic gap which yields minimum excitation. Therefore the design conditions are determined by the envelopes of the curves for excitation at each flux density. From the data for a large range of flux densities, design curves have been derived, Figs. 6–9. Each design curve for excitation permits the determination of excitation (the ordinate) and nonmagnetic gap from two independent quantities, flux density **B** (the abscissa) and magnetic field strength $\mathbf{H}_{DC}$ (the parameter). The appropriate values of effective gap are marked off on the curves. The design curves for finding core loss also yield this quantity as a function of density and magnetic field strength. The core loss values given are those for

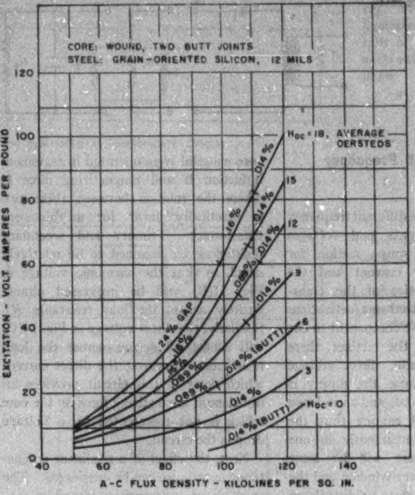


Fig. 6. Excitation and gap for wound core—design curves (60 cps)

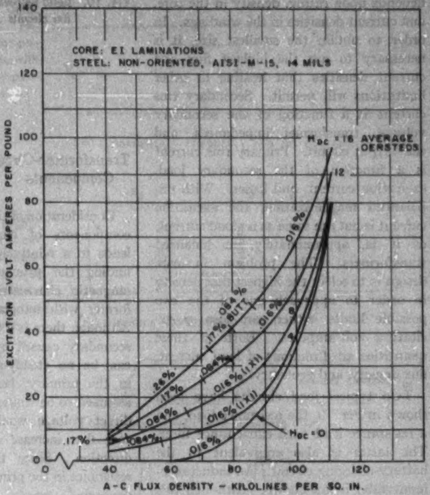


Fig. 8. Excitation and gap for stacked core—design curves (60 cps)

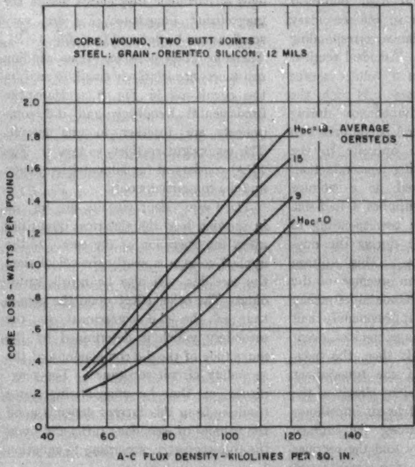


Fig. 7. Core loss of wound core—design curves (60 cps)

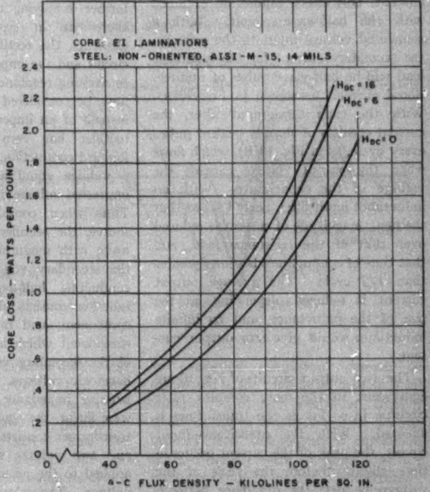


Fig. 9. Core loss of stacked core—design curves (60 cps)

the per-cent gap which yields minimum excitation. Qualitatively, it can be seen from the design curves that a nonmagnetic gap is likely to be indicated when flux density is intermediate or low, and field strength is high.

Half-Wave Rectifier Supply Transformer

The half-wave rectifier circuit, shown in Fig. 10, has been widely used, but has received little analysis. It is often used in certain important applications, such as bias power supplies, battery-charging circuits, and high-potential sources. The design or evaluation of transformers for half-wave rectifiers is similar to balanced types in that size depends upon rating, density in the core, and current densities in the windings. In order to obtain the smallest size, it is necessary to use the highest flux and current densities that heating or other limitations will permit. Secondary rms current is a function of the secondary voltage, transformer impedances, and the output circuit. Primary rms current is a function of the secondary load, excitation current, and losses. With unbalanced magnetization, the excitation current is not the same as no-load current, as it is, approximately, in balanced transformers. The problem in core design is to select the highest flux density in order to approach one of the two possible limits, either core loss or excitation volt-amperes. Both of these quantities are functions of load current, flux density, and geometry.

Two typical load circuits have been shown in Fig. 10, the resistance load, and a resistance load with capacitance filter. The latter is also equivalent to the battery-charging circuit. An inductance-input filter is not shown. Although this filter is used in other types of rectifiers to obtain an almost constant rectifier current, this effect cannot be obtained with the half-wave circuit. With a sinusoidal voltage input to the primary, the secondary current for the resistance load will be half-wave pulses of undirectional current, as shown in Fig. 10(A). With the capacitance-input filter, the current consists of sharply peaked pulses every cycle, as in Fig. 10(B), which flows when the output voltage exceeds the voltage of the capacitance. With an inductance-input filter, Seely[51] shows that the peak magnitude of current is reduced over that of the resistance load, and that time of conductance becomes greater than 1/2 cycle. The average output current is reduced, depending on the size of the inductance, and an infinite inductance would give zero output current.

The two output circuits of Fig. 10 are equivalent to the more complex filter circuits, in so far as the transformer is affected. With the capacitance-input circuit, the peak of the current occurs in time slightly before the peak of the voltage because the difference between secondary and condenser voltages is greater when secondary voltage is increased toward, rather than decreasing from the peak value. The ratio of secondary rms current to load direct current may be best found from curves given by Schade[50] as a function of circuit resistance, load resistance, and filter reactance.

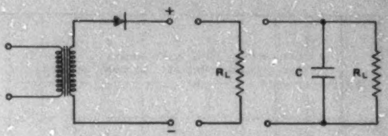


Fig. 10. Basic half-wave rectifier circuits

Transformer-Circuit Frequency Components

Consideration of the different frequency components of currents and voltages leads to a relatively simple method for finding the primary current and the magnetic characteristics of the transformer with unbalanced magnetization. Although there is a d-c component in the secondary caused by the rectifier, there can be no steady-state, direct voltage in the primary, because the supply is assumed to be sinusoidal, and an induced direct voltage would require that the core flux increase continuously in one direction. Since there will be some resistance in the primary winding and the connected supply, any unidirectional transient current will decay to zero.

In the circuit consisting of a transformer supplying a half-wave rectifier, there are in general two nonlinear elements, the rectifier in the secondary circuit and the impedance corresponding to exciting reactance. An ideal rectifier may be considered as a voltage source instead of an impedance. As such, the rectifier has zero impedance during forward conduction, and is equivalent to a voltage equal and opposite to the impressed voltage during nonconduction. Thus when considered as a voltage source, the rectifier supplies a half-sine wave with magnitude corresponding to the secondary voltage during the nonconducting half-cycles. This voltage function consists of an average of d-c component, and fundamental-frequency component which is approximately half of the secondary voltage, and of even-harmonic voltages. If, then, the magnetizing impedance of the transformer were linear, the equivalent circuit of the transformer consists of linear impedance and two voltage sources, the voltage applied to the primary and the rectifier voltage. The solution for currents and voltages can then be obtained from the individual equivalent circuits shown in Fig. 11. Fig. 11(A) is a circuit for fundamental-frequency currents; Fig. 11(B) is a circuit for harmonic frequencies; and Fig. 11(C) is a circuit for the d-c component.

The component circuits in Fig. 11(C) would be applicable in case the magnetic core material were operated in regions of induction **B** and magnetizing force **H** where the relation between these two is practically linear for an air-cored transformer. Primary and secondary resistances are assumed to be relatively small, so that the harmonic voltage in Fig. 11(B) will be impressed almost entirely across the load resistance R_L. Similarly, the direct voltage in Fig. 11(C) will almost all appear across the load resistance. However, the direct current which flows in this circuit provides an unbalanced magnetizing force on the core which is not represented by a voltage drop in the circuit.

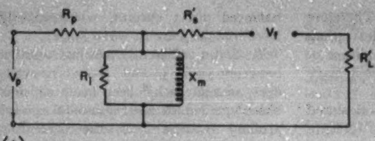
Next, the effect of a nonlinear magnetizing impedance can be considered. The principle of superposition prohibits separation of current and voltage frequency components associated with a nonlinear impedance. However, currents and voltages of harmonic frequencies across the magnetizing impedance X_m are very small, as can be deduced from Fig. 11(B). Therefore it is possible, when winding resistances are relatively small, to analyze the circuit as in Fig. 12. Here the fundamental frequency and d-c components are together in one circuit. The equivalent rectifier voltage in Fig. 12(A) consists of fundamental-frequency and d-c components only.

These equivalent circuits are an aid in showing how the electrical quantities affect the operation of the core. If the applied voltage is nearly sinusoidal then the core flux will also be nearly sinusoidal. The other effect upon the core is that of the d-c component in the secondary which is determined by the magnitude of the d-c component and the secondary circuit resistance. The magnitude of the d-c magnetizing force resulting from this current depends upon the number of secondary turns and upon the core geometry, according to equation 1.

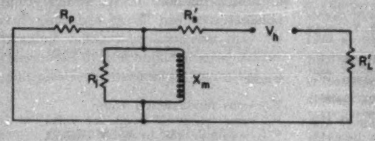
Primary Current

The significance of the equivalent-circuit analysis is that the alternating components of secondary ampere-turns are equal to the load component of primary ampere-turns. The sum of this load component and the excitation

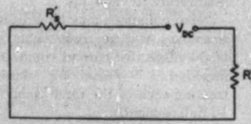
Fig. 11 (left). Equivalent circuits of component voltages with linear impedances



(A) FUNDAMENTAL-FREQUENCY CIRCUIT

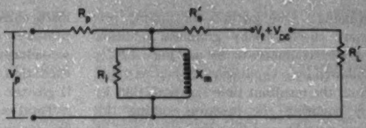


(B) HARMONIC-FREQUENCIES CIRCUIT

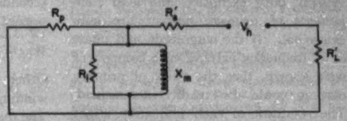


(C) D-C COMPONENT CIRCUIT

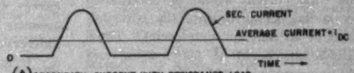
Fig. 12 (right). Equivalent circuits of component voltages with nonlinear magnetizing inductance



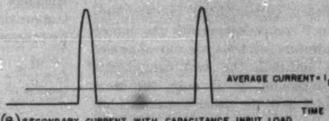
(A) FUNDAMENTAL-FREQUENCY AND DC COMPONENTS CIRCUIT



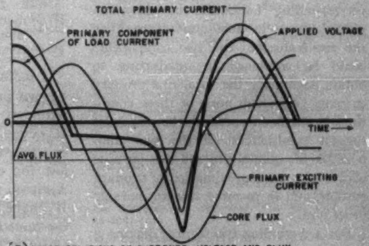
(B) HARMONIC-FREQUENCIES CIRCUIT



(A) SECONDARY CURRENT WITH RESISTANCE LOAD



(B) SECONDARY CURRENT WITH CAPACITANCE INPUT LOAD



(C) PHASE RELATIONS OF CURRENTS, VOLTAGE AND FLUX

Fig. 13. Wave shapes in half-wave rectifier supply transformer

component is the primary current. The a-c excitation-current function of time plus the secondary d-c component determines the total magnetomotive force which is related to the core flux function by the magnetization curve of the material. The primary component of load current for a half-wave rectifier with a resistance load is shown in Fig. 13(C) as a half sine-wave function without an average component. It is next necessary to establish the excitation component in terms of phase and qualitative form. Since the peak values of the flux function lag the peaks of the applied voltage by 1/4 cycle by definition, the peaks of the excitation component must also lag the voltage peaks by 1/4 cycle or 90 electrical degrees.

It must be found which peak of the excitation current is sharper because of the saturation effect. To do this, the primary load component of Fig. 13(C) can be considered for the moment as two currents, one exactly like the secondary current, including d-c component, plus a constant negative part equal to the average of the other. If this second negative part were absent, then average core flux would be zero, because the sum of the first part and secondary current would provide no unbalanced magnetomotive force. The fact that the hypothetical second part is negative indicates that saturation tends to occur during the negative half-cycles of the flux and excitation current, as shown in Fig. 13(C). This reasoning establishes the qualitative form of excitation current and the displacement of the average flux as given in the figure. Next, the sum of the load and excitation components gives the time function of total primary current. For very small flux densities the primary current becomes very similar to the given load component.

An ideal core material having infinite permeability cannot be assumed in an analysis such as this, because the presence of the unbalance in the secondary would yield an infinite flux. Therefore, an excitation component cannot be eliminated in the consideration of an ideal material. However, excitation current could be made small compared to the primary load component of current if the a-c density were so small that the cyclic variation in magnetic field strength (and corresponding ampere-turns) were relatively small compared to load-current ampere-turns. These conditions require a magnetic material with less than infinite permeability.

The qualitative shape of primary current has now been established, and next it is possible to find methods for calculating the rms magnitude which determines the size of wire needed for the winding. One method for finding primary current is by graphical addition of the components as in Fig. 13(C), followed by computation of the rms value from the resultant time function, that is, by squaring the function, taking the time average, and then the square root of the average. A better method is based upon elementary a-c circuit algebra. In general, both excitation and load components of primary current are nonsinusoidal. If the magnitude and phase of all frequency parts of each component were known, then the value of primary current could be readily determined. The resultant of each frequency might be found by combining the parts of the same frequency as phasors. Finally, the total current would be determined by adding in quadrature the contributions at all different frequencies. These calculations could be made using and obtaining rms values of current.

The excitation and load components of primary current have fundamental-frequency components which are almost in quadrature, or 1/4 cycle out of phase. If, in addition, the two components were sinusoidal, or if only one were sinusoidal and the other nonsinusoidal, or if the two components contained different higher harmonics, or if the corresponding harmonics in the components were in time quadrature, then the rms values of the two components could be combined in quadrature to obtain accurately the total rms primary current. Although none of these conditions is satisfied in general, the principles suggest combination of the two components in quadrature for the purpose of obtaining an approximation of primary current, and in practice, this is a suitable procedure.

For a 2-winding transformer having a turns ratio of unity, the rms load component of primary current is defined as the alternating parts of the secondary current function, which is simply the secondary current with the d-c component removed. Since the rms value of secondary current is precisely the quadrature sum of d-c and all a-c components, it follows that

$$I_{pL}=\sqrt{I_s^2-I_{DC}^2} \qquad (2)$$

where

I_{pL} = rms load component of primary current
I_s = rms secondary current
I_{DC} = average secondary current

For the case of an unfiltered output resistance (the first load circuit of Fig. 10), the secondary rms current is 1.57 I_{DC}, and the primary component of load current is 1.21 I_{DC}. Therefore primary current may be less than secondary current, but the addition of excitation in practical designs will make it greater.

Total primary input can be calculated by summing in-phase and quadrature volt-ampere components. The input component of output volt-amperes is secondary voltage times the load component of primary current.

$$W_{pL}=V_sI_{pL}=V_s\sqrt{I_s^2-I_{DC}^2} \qquad (3)$$

Other real power components are the winding losses W_c and core losses W_i. The magnetizing component of excitation power is $\sqrt{W_{ex}^2-W_i^2}$, where W_{ex} is the excitation volt-amperes given by the curves of the appendix. Leakage reactance is neglected here, so that the difference between primary and secondary terminal voltages (with unity turns ratio) is due to resistance drops in the windings. Approximate primary current is approximate input volt-amperes divided by primary voltage V_p.

$$I_p=\frac{W_{rp}}{V_p}=\frac{1}{V_p}\times \sqrt{(W_{pL}+W_c+W_i)^2+W_{ex}^2-W_i^2} \qquad (4)$$

where

W_{rp} = approximate total primary volt amperes
V_p = primary voltage

A more refined method than equation 4 is not justifiable in view of the variable nature of excitation among similar cores. However, the results obtained with the equation tend to be about 4 to 15% too low for electronic power transformers. A reason for this is that both W_{pL} and W_{ex} have higher harmonic components of the same frequency which are effectively added in quadrature regardless of actual phase. An inspection of Fig. 13(C) shows that the second harmonics of the two components are actually in phase (of the phase $+I_2 \cos 2\omega t$), so that the approximate calculation might be expected to give a low result.

Regulation and Turns Ratio

In transformers carrying sinusoidal currents, corrections for voltage drops can be made using the rms current and resistance for each winding (and the leakage reactance, if appreciable). When secondary current or voltage wave shapes differ appreciably from sinusoidal forms, the usual methods will yield voltage ratios which are excessively in error. This problem has been met in tests on model transformers with unbalanced direct current, which usually have complex current wave shapes.

To find a sufficiently accurate relation between primary and secondary voltages, an analysis has been made which is based upon an assumed sinusoidal applied primary voltage. If the secondary current is not sinusoidal, secondary voltage must likewise be nonsinusoidal. Secondary rms voltage may be defined as an exact function of its several components.

$$V_s=\sqrt{V_{sf}^2+V_{sh}^2+V_{sDC}^2} \text{ rms volts} \qquad (5)$$

where

V_{sf} = rms fundamental volts
V_{sh} = rms harmonic frequency volts
V_{sDC} = voltage due to direct current

The term V_{sh} is composed of the second and higher harmonics, all of which add in quadrature to yield V_{sh}.

Next, equations may be written for the secondary voltage components in terms of the respective current components and impedances. These are obtained by treating each of the three components of V_s independently.

$$V_{sf}=\frac{V_p}{n}-I_{sf}R' \qquad (6)$$

$$V_{sh}=I_{sh}R' \qquad (7)$$

$$V_{sDC}=I_{DC}R_s \qquad (8)$$

where

V_p = primary voltage
n = turns ratio
R' = transformer equivalent series resistance referred to the secondary
R_s = secondary resistance

To include the effects of leakage reactance, total impedance Z' could be substituted in equations 6 and 7 for R', and it would be necessary to consider the terms of the equations as phasors rather than as magnitudes. The calculation of secondary rms load voltage V_s from equations 5 to 8 can be considered as successive operations on the no-load secondary voltage V_p/n. First the term $I_{sf}R'$ is subtracted algebraically; then two other terms are added successively in quadrature. However, in transformers the voltage drop terms are always small compared to V_p/n. Only the term subtracted directly in equation 6, makes an appreciable change in V_p/n; the addition of the quadrature terms can be neglected. It follows that V_s is closely equal to V_{sf}, and that the difference between V_p/n and V_s depends primarily upon the voltage drop of the fundamental-frequency component of secondary current through the transformer equivalent resistance R'.

Summary

Magnetic circuits with unbalanced magnetization, in many respects, perform similarly to balanced circuits, but have additional characteristics which must be considered for accurate analysis or design. Special data on magnetic material properties are required to determine losses, excitation, and preferred gap of a core. To obtain these data, a convenient test circuit has been devised. Winding voltages and currents may be determined by the application of simple principles for nonsinusoidal quantities.

Appendix. Test Data

Pertinent data on typical transformer cores tested are:

1. *Wound-type core,* windings on one leg
   - Two butt joints
   - Mean length of magnetic circuit: 7.22 inches
   - Net cross-sectional area of core: 0.753 square inch
   - Core weight: 1.5 pounds
   - Steel: oriented silicon
   - Thickness: 12 mils
   - Test frequency: 60 cps (cycles per second)
2. *Shell-type core,* EI laminations (scrapless)
   - Leg width: 1 inch
   - Mean length of magnetic circuit: 6 inches
   - Core weight: 2.13 pounds
   - Steel grade: audio-type equivalent to *AISI-M-15,* nonoriented silicon
   - Thickness: 14 mils
   - Test frequency: 60 cps

The following equations give the approximate per-cent gap which results in minimum excitation. In each of the cores it has been found that the per-cent gap is almost independent of flux density below some value of flux density.

1. Per-cent gap $=0.015\ \mathbf{H}_{DC}-0.003\ \mathbf{B}+0.24$
   (Use $\mathbf{B}=90$ if density is less than 90),
   where $\mathbf{H}_{DC}$ = magnetic field strength in oersteds, 0.495 times average ampere-turns per inch
   $\mathbf{B}$ = flux density in kilolines per square inch
2. Per-cent gap $=0.021\ \mathbf{H}_{DC}-0.016\ \mathbf{B}+1.0$
   (Use $\mathbf{B}=70$ if density is less than 70)

Application of these equations might yield a negative value for the per-cent gap. In this case the core joint with minimum effective gap should be used, a butt joint for cut, wound cores, and a lapped-butt joint for stacked cores.

References

1. Bibliography on Transducers, Magnetic Amplifiers, **H. B. Rex.** *Instruments,* vol. 21, Apr. 1948, p. 332.

2. Bibliography of Magnetic Amplifier Devices and the Saturable Reactor Art, **James G. Miles.** *AIEE Transactions,* vol. 70, pt. II, 1951, pp. 2104–23.

3. Hysteresis in Iron for Asymmetric Cycles of Magnetization, **G. Vallauri.** *Atti Dello Association Elettrotecnica Italiana,* vol. 15, 1911, pp. 79–92.

4. Hysteresis Loss in Iron Taken Through Unsymmetrical Cycles of Constant Amplitude, **M. Rosenbaum.** *Journal,* Institution of Electrical Engineers, vol. 48, 1912, pp. 534–45.

5. Experiments on Magnetic Hysteresis, **F. Holm.** *Elektrotechnische Zeitschrift,* vol. 33, 1912, pp. 928–29.

6. Effect of Displaced Magnetic Pulsations on the Hysteresis Loss of Sheet Steel, **L. W. Chubb, T. Spooner.** *AIEE Proceedings,* vol. 34, 1915, pp. 2321–42.

7. The Unsymmetrical Hysteresis Loop, **John D. Ball.** *AIEE Transactions,* vol. 34, pt. II, 1915, pp. 2693–20.

8. The Alternating Flux in an Iron Core with Superposed Direct Current Magnetization, **H. Schunck.** *Archiv für Elektrotechnik,* vol. 12, 1923, pp. 428–33.

9. Magnetic Properties of Sheet Steel Under Superposed Alternating Field and Unsymmetrical Hysteresis Losses, **Y. Niwa, Y. Asami.** *Researches of the Electrotechnical Laboratory,* No. 124, June 1923.

10. Further Study of the Magnetic Properties of Electrical Sheet Steel Under a Superposed Alternating Field and Unsymmetrical Hysteresis Losses, **Y. Niwa, J. Sugiura, J. Matura.** *Ibid.,* No. 144, May 1924.

11. Effect of a Superposed Alternating Field on Apparent Magnetic Permeability and Hysteresis Loss, **T. Spooner.** *Physical Review,* vol. 25, 1925, pp. 527–40.

12. Incremental Magnetization, **L. G. A. Sims, D. L. Clay.** *Wireless Engineer,* vol. 12, May 1935, pp. 238–45; June, 1935, pp. 312–20.

13. Some Data Concerning Superposed D.C. and A.C. Magnetization of Modern Transformer Sheets, **H. G. Nolen, J. G. Niesten.** *Electrotechnik,* vol. 24, Apr. 11, 1946, pp. 85–89; Apr. 25, 1946, pp. 103–06; May 9, 1946, pp. 114–15.

14. On the Variations of A.C. Permeability of Transformer Sheet Steels with D.C. Magnetization, **B. M. Banerjee.** *Indian Journal of Physics,* vol. 22, 1948, pp. 265–75.

15. Research and Development of Various Configurations of Core Materials for Optimum Transformer Design, Final Report. Battelle Memorial Institute, Columbus, Ohio, *Contract No. W 36-039-sc-38255,* Department of the Army, Signal Corps Engineering Laboratories, 1952.

16. Research and Development of New Design Method for Power Transformers, Final Report. Armour Research Foundation, Chicago, Ill., *Contract No. DA 36-039-sc-52656,* Department of the Army, Signal Corps Engineering Laboratories, 1955.

17. Bridge Circuit for Measurement of Inductance Coils with A.C. Superimposed on D.C., **V. D. Landon.** *Proceedings,* Institute of Radio Engineers, vol. 16, 1928, pp. 1771–75.

18. Improved Form of Maxwell D.C. Inductance Bridge, **L. H. Harris, H. Williams.** *Post Office Electrical Engineers Journal,* vol. 23, 1930, pp. 36–41.

19. Inductance Measurements, **R. D. Salmon.** *Electrician,* vol. 113, 1934, pp. 827–29.

20. Measuring Inductance of Coils with Superimposed Direct Current, **H. T. Wilhelm.** *Bell Laboratories Record,* vol. 14, 1935, pp. 131–35.

21. A.C. Impedance of Chokes and Transformers, **T. J. Rehfisch, H. T. Bissmire.** *Wireless Engineer,* vol. 18, 1941, pp. 266–70.

22. The Theory and Design of Inductance Coils (book), **V. G. Welsby.** MacDonald and Company, London, England, 1950.

23. Electrical Measurements (book), **F. K. Harris.** John Wiley & Sons, Inc., New York, N. Y., 1952.

24. A Wide Range A-C Bridge Test for Magnetic Materials, **D. C. Dieterly, C. E. Ward.** *Bulletin,* American Society for Testing Materials, May 1952, pp. 75–80.

25. Ferroinductance as a Variable Electric-Circuit Element, **J. D. Ryder.** *AIEE Transactions,* vol. 64, Oct. 1945, pp. 671–78.

26. Losses in Iron Under the Action of Superposed Alternating- and Direct-Current Excitations, **O. E. Charlton, J. E. Jackson.** *Ibid.,* vol. 44, 1925, pp. 824–31.

27. Effective Inductance, **H. Teuchert.** *Elektrotechnische Zeitschrift,* vol. 53, 1932, pp. 103–04.

28. Incremental Permeability and Inductance. **L. G. A. Sims.** *Wireless Engineer,* vol. 12, 1935, pp. 8–16, 65–69.

29. Measurement of Inductance of D.C. Loaded Transformers, **W. Nowotny, F. Söchting.** *Elektrotechnik und Maschinenbau,* vol. 54, 1936, pp. 473–74.

30. Reactor Measurements, **H. L. Daniels.** *Electronics,* vol. 18, 1945, pp. 146–48.

31. A Method of Using Transformers as Choking Coils, **J. D. Coales.** *Journal,* Institution of Electrical Engineers, vol. 42, 1909, pp. 412–54.

32. Design of Reactances and Transformers which Carry Direct Current, **C. R. Hanna.** *AIEE Transactions,* vol. 46, 1927, pp. 155–60.

33. Design of Choke Coils and Transformers which Carry a Direct Current, **G. W. O. Howe.** *Wireless Engineer,* vol. 5, 1928, pp. 49–52.

34. Design Methods for Soft Magnetic Materials in Radio, **J. Minton, I. G. Maloff.** *Proceedings,* Institute of Radio Engineers, vol. 17, 1929, pp. 1021–33.

35. Constant Impedance Method for Measuring Inductance of Choke Coils, **F. M. Turner.** *Ibid.,* vol. 16, 1928, pp. 1559–70.

36. Inductance as Affected by the Initial Magnetic State, Air-Gap and Superimposed Currents, **F. M. Turner.** *Ibid.,* vol. 17, 1929, pp. 1822–33.

37. Polarized Magnetic Circuits, **R. Gürtler.** *Hochfrequenztechnik und Elektroakustik,* vol. 39, Jan. 1932, pp. 2–7; Feb. 1932, pp. 48–59.

38. Economy of Polarized Inductances, **R. Gürtler.** *Ibid.,* vol. 39, May 1932, pp. 171–73.

39. The Air Gap Transformer and Choke, **F. W. Lanchester.** *Journal,* Institution of Electrical Engineers, vol. 73, 1933, pp. 413–18.

40. Alternating-Current Inductance of an Iron-Cored Coil Carrying DC, **R. T. Beatty.** *Wireless Engineer,* vol. 11, 1934, pp. 61–63.

41. Superimposed DC and AC in Iron-Cored Transformers and Chokes, **C. R. Cosens.** *Ibid.,* vol. 12, 1935, pp. 190–93.

42. Iron Cored Inductances, **W. Baggally.** *Ibid.,* vol. 13, 1936, pp. 7–9.

43. The Inductance of Iron-Cored Coils Carrying Direct Current, **G. F. Partridge.** *The Philosophical Magazine,* vol. 23, 1937, pp. 99–106.

44. The Inductor with Air-Gapped Magnetic Circuit, **E. V. D. Glazier.** *Engineering,* vol. 148, 1939, pp. 406–08.

45. Reactors in D.C. Service, **R. Lee.** *Electronics,* vol. 9, 1936, pp. 18–20, 76.

46. Electronic Transformers and Circuits (book), **R. Lee.** John Wiley & Sons, Inc., New York, N. Y., 1947.

47. Optimum Air Gap for Various Magnetic Materials in Cores of Coils Subject to Superposed Direct Current, **V. E. Legg.** *AIEE Transactions,* vol. 64, Oct. 1945, pp. 709–12.

48. Incremental Magnetic Properties of Silicon Steel, With Particular Reference to the Design of Air-Gapped Smoothing Chokes, **R. O. Carter, D. L. Richards.** *Proceedings,* Institution of Electrical Engineers, vol. 97, 1950, pp. 199–214.

49. Design of Iron-Cored Inductances Carrying D.C., **N. H. Crowhurst.** *Electronic Engineering,* vol. 22, 1950, pp. 516–23.

50. Analysis of Rectifier Operation, **O. H. Schade.** *Proceedings,* Institute of Radio Engineers, vol. 31, 1943, pp. 341–61.

51. Electron-Tube Circuits (book), **Samuel Seely.** McGraw-Hill Book Company, Inc., New York, N. Y., 1950.

52. The Transformer for Measuring Large

Direct Currents, Harris J. Ryan. *AIEE Transactions*, vol. 18, 1901, pp. 169–90.

53. Measurement of Large Direct Currents Over Long Distances, E. Besag. *Elektrotechnische Zeitschrift*, vol. 40, 1919, pp. 436–37.

54. Simple D.C. Instrument Transformers with True Current Transformer Properties, W. Krämer. *Ibid.*, vol. 58, 1937, pp. 1309–13.

55. The Theory of the Current Transductor and Its Application in the Aluminum Industry, T. R. Specht, R. N. Wagner. *AIEE Transactions*, vol. 69, pt. I, 1950, pp. 441–52.

56. Theory of D-C Excited Iron-Core Reactors and Regulators, A. Boyajian. *Ibid.*, vol. 43, 1924, pp. 919–36.

57. Universal Curves for D-C Controllable Reactors, Walter C. Johnson, B. C. Merrell, R. E. Alley, Jr. *Ibid.*, vol. 68, pt. I, 1949, pp. 31–40.

58. The Magnetic Amplifier as a Nonlinear Circuit Element, E. J. Smith. *Proceedings*, Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, Apr. 23–24, 1953, pp. 274–319.

59. Magnetic Circuits with D-C Magnetization and High Alternating Induction, H. L. Garbarino. *Ph.D. Thesis*, Illinois Institute of Technology, Chicago, Ill., June 1955.

List of Periodicals

AIEE Proceedings, New York, N. Y.

AIEE Transactions, New York, N. Y.

Archiv für Elektrotechnik, Berlin-Charlottenburg, Germany

Atti Della Associazione Elettrotecnica Italiana, Milan, Italy

Bell Laboratories Record, New York, N.Y.

Bulletin, American Society for Testing Materials, Philadelphia, Pa.

Electrician, London, England

Electronic Engineering, London, England

Electronics, New York, N. Y.

Elektrotechnik, Berlin, Germany

Elektrotechnik und Machinenbau, Vienna, Austria

Elektrotechnische Zeitschrift, Wuppertal-Elberfeld, Germany

Engineering, London, England

Hochfrequenztechnik und Elektroakustik, Leipzig, Germany

Indian Journal of Physics, Calcutta, India

Instruments, Pittsburgh, Pa.

Journal, Institution of Electrical Engineers, London, England

The Philosophical Magazine, London, England

Physical Review, New York, N. Y.

Post Office Electrical Engineers Journal, London, England

Proceedings, Institute of Radio Engineers, New York, N. Y.

Proceedings, Institution of Electrical Engineers, London, England

Proceedings, Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

Researches of the Electrotechnical Laboratories, Tokyo, Japan

Wireless Engineer, London, England

New Concepts in Evaluating Electric Equipment Insulation Systems—The Problems Encountered

M. L. MANNING
FELLOW AIEE

ENGINEERS have searched for many years to find suitable yardsticks to evaluate insulation. On a scientific plane, one of the earliest approaches was made by Lamme and Steinmetz in 1912.[1] Ultimately, their work resulted in the preparation of AIEE Standard no. *1*.[2] This Standard outlines the principles underlying the accepted method of classifying electrical insulation material. As long as reliance was placed on the insulating materials that existed in nature, it was reasonable to classify materials on the basis of their chemical composition. But, within the past few years, the chemical industry has made great advances. Today, approximately 350 bonding and impregnating resins known by various trade names are available. With the coming of these chemical investigations and the creation of truly synthetic materials, it is no longer rational to fit existing insulations into the old chemical-composition classes.

Experimental work by users and manufacturers alike of electric equipment has disclosed the need for an entirely new concept of insulation classification and evaluation. It was found that insulations should be classified as systems as well as materials. The manner in which the materials are applied and processed is of equal importance with the basic materials used. The electrical industry, working through the AIEE, developed the concept of functional evaluation of insulation systems. This concept now appears in the revised AIEE Standard no. *1* published in 1954.[3] Revisions are scheduled for publication about June 1957.

What does the philosophy mean to users of electric equipment?

1. It recognizes the determination of insulation life by tests made in comparison with an existing insulation system in a motor or in a transformer.
2. It necessitates the re-examination of all insulation systems used in electric machinery by the manufacturer.
3. It provides assurance that the manufacturer has taken care to provide insulation systems proved by prescribed test procedures.
4. It gives more information concerning end of life of electric equipment using certain insulation systems.

Performance of insulating materials is of utmost concern to anyone using or manufacturing electric equipment. Most troubles of electric devices can be traced to limitations of insulating materials, to their inability to withstand temperature, atmospheric or environmental conditions, voltage disturbances, short circuits and overloads. Conventional books on the subject fail to keep pace with the swift progress of research, development, and applications. Courses of study in technical schools have not recognized sufficiently the importance of the subject.

This paper will outline why there is a need for new concepts; a brief review of the kinds of materials generally used in electrical equipment in fluid, solid, and gaseous form; the use of solid insulation in dry-type transformers and other equipment rated 80 C (degrees centigrade) and 150 C temperature rise by resistance respectively; application of gases and bibliography of AIEE papers soon to be available; complexities found in the evaluation of insulation and unanswered questions to be faced today; basic factors to consider in determining the hottest-spot temperature in windings of transformers in particular; basic considerations to be faced in the preparation of Standards for appraising or evaluating

---

Paper 57-392, recommended by the AIEE Dielectrics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Southern District Meeting, Jackson, Miss., April 3–5, 1957. Manuscript submitted January 7, 1957; made available for printing February 8, 1957.

M. L. Manning is with the McGraw-Edison Company, Canonsburg, Pa.

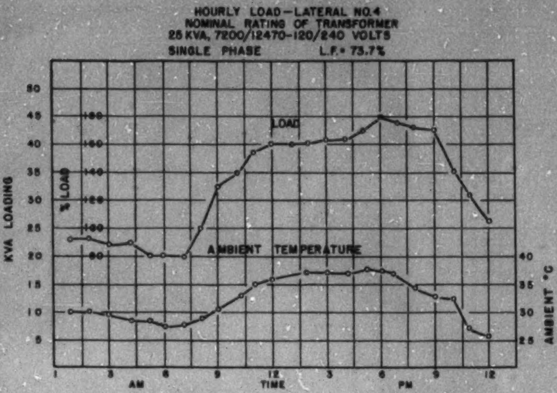


Fig. 1. Load curve for a Southwest power company: residential load study. Peak September 20, 1954

the insulation life of equipment; what the AIEE is doing about these problems; what manufacturers are doing; and lastly, of great importance, what users of equipment can do to assist in this complex picture.

The Need for New Concepts

Advances in the chemical industry have produced quite amazing progress in creating new bonding and impregnating resins and elastomers, new glass and asbestos base, and mica structural materials. These include solventless, hardenable casting and impregnating resins, solid 3-dimensional thermosetting insulating materials, insulating varnishes containing solvents dried by chemical processes, solid laminated insulating material, films and sheets, hot-bonding laminated materials, and solid thermoplastic nonlaminated insulating materials.[4]

Rapid electric-power company expansion has forced attention to overloaded electrical equipment. Load growths are doubling in 4 to 10 years. Air-conditioning loading continues to multiply, particularly in the Southwest. Distribution demands increase, more kva per pole is desired. Expansion programs of industry, new buildings, and housing developments create unprecedented blocks of power consumption. "Crowd your electrical equipment in small space," is the cry of contractors.

Coupled with small space requirements, unusual "heat storms" may occur in certain areas of the country. It is not uncommon for power companies in the Southwest to experience normal loading at 7 a.m., followed by 160% loading at noon continuously until 5 p.m., to 180% load at 6 p.m., dropping to 100% load by midnight; see Figs. 1 and 2. Such loading imposes abnormal temperature conditions under high ambient temperatures on connected apparatus. Normal permissible hottest-spot temperatures are exceeded for electric equipment, particularly transformers. How much life is taken from insulation systems under these accelerated conditions? What amortization value can be used: 4 years, 7 years, 10 years, 30 years? These questions compel consideration of the new concept of evaluating insulation systems under accelerated conditions with respect to temperature cycling, voltage, environmental atmospheres, and mechanical shock. By extrapolation from several cycling temperatures, life can be estimated at normal hottest-spot temperatures. It must be remembered that such an estimate is made under accelerated conditions. A statistical analysis is made here of test data to formulate answers.

To realize the situation more clearly for oil-filled transformers, basic data published over the past 25 years may be plotted on reciprocal absolute temperature scale graph paper; see Fig. 3. The figure will show that authors are, for example, in disagreement about the life at 100 C. In considering the points of curve 3, Fig. 3 (see references 5 through 8), which refer to the 1956 standard, "Guide for Loading Oil-Immersed Distribution and Power Transformers,"[7] the reader must realize that an attempt was made by the author to determine over-all loss of life. Perhaps Fig. 92.840 of the 1956 standard was not intended to be used exactly in this manner.[9] It can be contrasted with Appendix I, Fig. 4, of the 1948 edition. If it is attempted to make some kind of an estimate in spite of the fact that the 1956 standard (section 92-843) states, "The transformer committee does not recommend that any particular transformer life expectancy be assumed," the following tabulation may be given as an approximation. It is to be used with the understanding that the values do not appear in the standards under conditions assumed and the form tabulated. As an example of approximation, consider 100 C. From Fig. 92.840, loss of life per 24 hours is 0.132%. Expected life is

$$\frac{100}{0.132} = 758 \text{ days} = 2.08 \text{ years}$$

Hottest-Spot Temperature, C	ASA C57.92 (1956) Fig. 92.840 Life (Years) Estimated from 100% Basis, Approximation Only
97	5.4
100	2.08
110	0.48

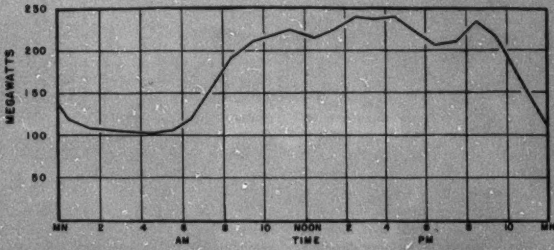


Fig. 2. Southwest power company system load curve: maximum temperature 35 C; minimum 25 C; May 30, 1956

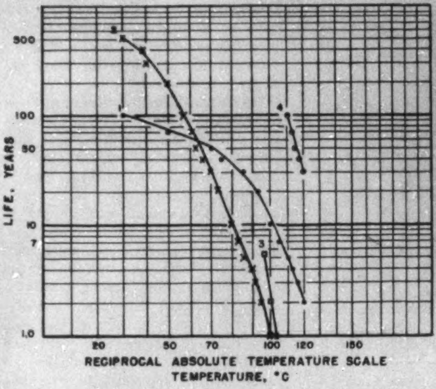


Fig. 3. Aging data: oil-filled transformers. See references as follows: curve 1, reference 5; curve 2, reference 6; curve 3, reference 7; and curve 4, reference 8

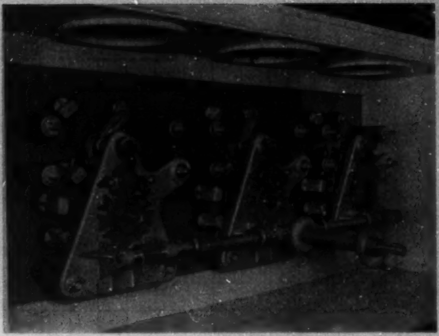


Fig. 4. Low-loss laminate LTC panel

The reader may infer from Fig. 3 that in spite of all the knowledge available today, as many answers seem to be offered for life expectancy as there are authors who treat the subject. This means that committees have much work to do in the future. At present, the user of electric equipment must decide which of the published data will apply to his needs.

Kinds of Insulating Materials Generally Used in Electric Equipment

Basically, insulating materials may be classified as fluids, solids, and gases.

FLUIDS

Oil

Heat is a limiting factor in loading of equipment. Oil picks up heat from immersed windings and transfers it to cooling surfaces. Various methods of oil preservation, nitrogen filling of the space above the oil, and the use of inhibitors are factors to consider in evaluating life of oil. Usually oil is limited to 100 C for protected systems. The viscosity of transformer oil in particular decreases rapidly with temperature increase. This leads to increased oil flow with increased loads. The decrease in viscosity with temperature increase results in lower differences of temperature between windings and top oil at high temperatures. This is fortunate since hottest-spot temperatures are reduced. It is quite interesting to observe that roughly only 0.3 pint of oil is used per kva in modern forced-oil-cooled 250,000-kva transformers today. Therefore constant attention is paid to the properties of oil, since its cost is low for its importance.

Miscellaneous Fluids

Of recent development, the silicone fluids are found to have interesting properties which make them suitable for high temperatures up to 250 C.[10] The use of these fluids will be watched with interest.

SOLID INSULATIONS: OIL-FILLED TRANSFORMERS AND OTHER 105 C SYSTEMS

In the field of laminated materials, are found the following:

Masonite

Exploded wood fibers bonded by phenolic and other types of resins built up from 1/4-inch through 2-inch thicknesses. It is used in structural parts of transformers such as axial spacers, core blocking, etc., usually through 46-kv classes.

Pressboard

Presently, considerable discussion exists about blends to be used, i.e., 50% cotton rag and 50% kraft, or 75% cotton and 25% kraft, or 100% kraft. Arguments are advanced with respect to densities having a range of values from 0.95 to 1.1 and over. It is claimed that more oil penetration is allowed for lower densities.

Photographs of 75-25 blend pressboard at 90 times magnification are interesting. It is important to remember that collars and channels require severe bending stress characteristics. A fibrous cotton structure, it would seem, is helpful. Pressboard of 50-50 composition shows a dielectric strength about 30% lower than 75-25 blend using 2-inch diameter electrodes in oil.

Paper Base Laminate: Low Loss

A low-loss paper base laminate has an interesting record of experience since its development in 1950. It also has an outstanding record for use in load tap-changing (LTC) terminal boards and in selector switch panels. The average 60-cycle power factor value as received at 25 C is 2% or less. With the use of 1/2-inch diameter pins spaced 3$^1/_2$ inches apart, the withstand 60-cycle voltage, 1-minute strength, is 100 kv. This illustrates the interlaminar strength. Impulse voltages up to 335 kv have been withstood for the spacing mentioned. Since tracking in solid insulation often raises a question to be answered, a panel of this low-loss laminate was placed in heavily carbonized oil for a period of 24 days, and 50 kv applied for the pin spacing given. The dielectric strength of the carbonized oil was 17.5 kv. No failure was experienced. Under impulse voltage tests in a period of a few microseconds, carbon was precipitated in stress patterns at 300 kv. Flashover occurred through the oil but not on the surface of the laminate. About 600 panels of this low-loss (low power factor) laminate in large sheet sizes in 1/2- to 1-inch thicknesses are in service without failure. In the industry, prior to this development, large transformers were subject to failure of the LTC equipment panels. An important criterion in development work was the power factor method whereby power factor versus voltage relation is obtained. This test detects the presence of voids and verifies the use of proper resins and procedures.

And last but not least, the importance of raised electrodes on terminal boards must be realized. Voltage stress can be equalized in the oil space by use of electrodes mounted above the low-loss terminal board such as that shown in Fig. 4.

Another recent and important development is rolled paper base low-loss resin-bonded tubing for drive shafts in LTC equipment and for tap changer shafts in power and distribution transformers. This tubing possesses a preferable power factor of 4% 60-cycle voltage at room temperature. It has a tensile strength of 6,000 pounds per square inch and 13,000 pounds per square inch minimum in compression.

Wood Laminates for Bolts

These bolts are used for the high-voltage lead-bracing construction of power transformers. The laminate is composed of a phenolic impregnated, compressed lamina of wood plies. The impact strength is high, up to 7 foot-pounds per inch on the Izod test and flexural strength of 35,000 pounds per square inch. Use of this material assists in obtaining reduced clearances to ground.

Kraft Paper

In distribution transformers, a trend is elimination of varnish impregnation of coils. High on the list of usable layer insulation stands shellac-coated kraft paper.

This 100% base sulfate wood pulp fiber paper has an ash content not greater than 1.0%, maximum moisture content as received 7%, and density from 0.95 to 1.0, depending on thickness.

Shellac produced by a lac insect contains a mixture of polyhydroxy acids and esters. On drying, shellac esters are formed which causes the molecular weight to build up considerably. This makes less sensitivity to moisture and to solvents than more simple esters. No man in a chemical laboratory as yet has succeeded in duplicating the properties of shellac as nature has formulated it. Of great importance is the method used to put shellac on the paper. An initial cure at 170 Fahrenheit is usually given. Layer insulation and pressboard spacers, shellac-coated, provide bond strength when given a final cure at 140 C (285 Fahrenheit). Under oil, after the cure temperature mentioned, the dielectric strength of 10-mil shellac-coated kraft paper approximates 700 VPM, using 2-inch diameter electrodes.

Moisture sensitivity of insulating materials having organic fibrous structures, such as kraft or cotton, is often overlooked. In 90% relative humidity atmosphere, 10-mil untreated bone-dry kraft paper picks up an 11% increase in weight within 24 hours. As much increase in weight is obtained within the first 24 hours as in succeeding 300-hour periods. This emphasizes the importance of humidity control in storage conditions.

Manila Rope Paper For Wire Covering

Manila rope paper possesses the tensile strength required for bending and shaping wire into coils. And a review of technical papers related to aging tests on paper shows that most of the work has been done on this kind of material. This paper contains 100% rope fibers, and on standard American Society for Testing Materials tests a dielectric strength of 150 volts per mil for untreated paper is found. Tensile strength varies from 25 pounds per inch for a 1-mil thickness to 42 pounds per inch for 3-mil thicknesses.

Formvar Covered Wire, Heavy Coating

Formvar resin-covered wire is used quite extensively in round-wire distribution transformer windings. It is resistant to solvents and to scrape tests. This wire possesses high flexural strength and is quite elastic. Formvar or polyvinyl formal has unique properties due to its tough elastomeric composition. It has been used in the lining of fruit cans and in self-sealing gasoline tanks. Scrape resistance is 250% improved over conventional enamels.

Gases, the third type of insulating material, are covered in a later section.

Dry-Type Transformers and Other Equipment Rated 80 C or 150 C Rise: 150 C or 220 C Permissible Hottest-Spot Temperatures

AIEE committees, in particular, are busy in work related to screening basic materials into letter classifications with hottest-spot limits and without defining the materials with respect to composition. From this classification, having a test procedure, manufacturers will incorporate the materials into insulation systems evaluated functionally by test. National Electrical Manufacturers Association transformer standards now classify dry-type transformers into three groups, 1, 2, and 3, having permissible hottest-spot temperatures of 105, 150, and 220 C respectively.[11]

The AIEE working group on life of materials committee has submitted for approval a "Proposed Test Procedure for Functional Temperature Endurance Tests of Ventilated Dry-Type and Distribution Transformers." This test procedure establishes procedures for evaluating insulating systems designed for such transformers.[12]

When speaking of insulation systems for dry-type transformers, one cannot overlook the new materials known as silicones or "magic sand." By use of these silicone resins and rubbers in transformers, the operating temperatures can be increased an additional 70 C over previously used insulation for dry-type units. Moreover, freedom from fire and explosion, due to the molecular construction of the materials, is assured. These ventilated or sealed-in-nitrogen transformers built in ratings through 2,000 kva, 15 kv, are the safest ever manufactured for unit substation, station auxiliary, and network application.

A description test based on functional evaluation procedures outlined by the AIEE working group on life of materials for a silicone system was published in April 1955.[13] Other insulation systems are considered in papers appearing in 1955 and 1956.[14–16]

Varied Applications in Electric Equipment

Application of solid insulating materials and the complexities of the many forms can be judged by noting the efforts of the International Electrotechnical Commission in writing an Encyclopedia of Electrical Insulation.[4] In its classifications lies the possibility of distinguishing between 40 groups of insulating materials. Each group is classified as:

1. Final state.
2. Function.
3. State when applied.
4. Unique problems in applying, such as use of a solvent.

Laminated materials are outlined with respect to bonding resins such as the phenolics, melamines, polyesters, silicones and teflon. Films and sheets are listed as cellulose, polyethylene, polystyrene, polyvinyl chloride, polyamide, and rubbers. Hot bonding laminated and nonlaminated materials are arranged in detail. The complexity of using known materials today demands the concept of evaluating insulation systems on a functional and temperature basis.

In other applications of concepts described for functional or prototype evaluation, several AIEE papers have been written for the use of small dry-type transformers in electronic equipment[17] and for motors.[18–21] The AIEE publication "Proposed Test Procedure for Evaluation of the Thermal Stability of Enameled Wire" is now available.[22] References to papers listed in this test procedure are valuable.

Gases, Applications

A library survey of gaseous dielectric phenomena is quite enlightening. The AIEE subcommittee under the committee of dielectrics has prepared a bibliography of papers dealing with spark breakdown, corona phenomena, practical considerations, theoretical and didactic, physical and chemical aspects. This bibliography, to be published in 1957, will be of considerable value. On practical applications alone, at least 50 papers have been written from 1925 to 1955 inclusive. Recent papers list properties of nitrogen compared with properties of sulphur hexafluoride and perfluorocarbon gases.[23,24]

Thermal conductivity and the electric strength of some gases are interesting. In fact, under special conditions the electric strengths of some gases are superior to those of solid and liquid insulations. Fluorogases of most interest are the Freons, sulphur hexafluoride and the perfluorocarbon gases possessing various molecular weights up to 624 for $C_{14}F_{24}$ (perfluorophenanthrane). Depending upon application, special care must be exercised in choosing a gas having the boiling point desired, freedom from toxicity, effect of the gas upon the insulation system and electric strength of the gas at ambient as well as at operating temperatures. Perfluoropropane C_3F_8 is of interest in application to sealed dry-type transformers. This colorless gas is nonflammable, nontoxic, has a boiling point of −38 C, and is practically the same in thermal conductivity properties as transformer oil. Tests reveal that 500-kva, sealed-in-C_3F_8-gas, dry-type silicone insulated transformers can carry 20% overloads for the same temperature rise as is obtainable for nitrogen-filled units at 100% load. This gas will not soften silicone resins and elastomers.

It must be remembered also that insulation systems sealed in nitrogen and other gaseous atmospheres described possess, as a rule, better aging characteristics than exposure to air. This fact must be considered in determining the life of a system.[25]

Complexities Found in Evaluation of Insulation

Temperature is not the only limitation imposed on an insulation system. Endurance under mechanical and electrical stress, vibration, environmental conditions and contaminants play predominant parts.

A few highlights to consider in choosing temperature limits are as follows:

1. Effects of long exposure to high temperature. Do the materials change with respect to mechanical and electrical properties? Do the materials soften, lose weight, or become brittle?

2. The magnitude and time duration of loading imposed on the insulation system.

3. Physical support. What happens to the physical support of insulating materials at high temperatures? The bonding resins may become brittle, yet the insulation system may continue to function if not disturbed. Forces of thermal expansion and contraction may restrict temperature limitations depending upon the size of electrical equipment.

4. Life of insulation. Insulation life is dependent upon environment such as moisture, dirt, chemicals or solvents released from the interior of an insulation system. Various industrial atmospheres may have predominant influence on ventilated apparatus compared with sealed systems which exclude oxidation or contamination.

5. Deterioration. The rate of physical deterioration increases rapidly as temperature increases. The modern concept of determining insulation life at elevated temperature is based upon the idea that the logarithm of insulation life is a function of the reciprocal of the absolute temperature. At three or more temperatures the logarithm of the hours of life found by thermal evaluation tests are plotted against the reciprocal of the absolute temperatures. Usually a straight line results. The life at other temperatures can be estimated. Individual time-temperature life curves for different materials or insulation systems can be determined by thermal evaluation tests.[26]

$$K = A\epsilon^{\frac{-E}{RT}}$$

where

K = specific reaction rate
E = activation energy
T = absolute temperature
R = universal gas constant
A = frequency of molecular encounter

This equation may be expressed in logarithmic form by the following steps:

Let

$$\frac{E}{R} = B$$

$$T = 273\text{ C} + \theta$$

$$\epsilon = 2.718$$

$$K = L \text{ (life)}$$

Then

$$\ln L = \frac{B}{T}\ln(2.718) + \ln A$$

$$= \frac{B}{\theta + 273} + \ln A$$

This equation can be recognized as the form

$$Y = a + bx \qquad (1)$$

by which can be written by substitution

$$Y = \text{insulation life} \sim \frac{1}{\text{reaction rate}} = A\epsilon^{\frac{B}{T}}$$

$$\ln Y = \ln A + \frac{B}{T} \qquad (2)$$

which may be compared with equation 1

when $Y = \ln y$
$a = \ln A$
$x = \frac{1}{T}$

The problem is to use equation 1 and then find values of constants a and b to give a straight line. Such is the theory behind reciprocal absolute temperature scale graph paper on which aging data are plotted.

6. Ambient temperature. Ambient temperature affects the temperature attained by equipment in operation.

7. Insulation life. Insulation life depends on care and maintenance of equipment in which it is used. Mechanical shocks, such as movement of equipment to different locations, short circuits, overvoltage are all factors to consider.

In general, a method is needed which, like a mill, would grind out functional merit ratings of insulation systems. All of the factors considered may be likened to seven prongs impinging upon insulation simultaneously.

A factor often overlooked is that a void-free structure is essential if the insulation is to be effective in dissipating heat. It must be compact, and free of cracks or air pockets. Voids are a source of corona formation.

Some unanswered questions are:

1. What causes thermal degradation? Are all demands taken care of, for example, in evaluating laminates for flexibility, hardness, and thermal stability?

2. What are the effects of insulation thickness upon aging properties (10-mil to 100-mil thickness) in air or in various gases?

3. Is a course being taken to supply distinct types of varnishes for specific applications for motors only? for transformers only?

4. What end points are involved—5 years? 10 years? 25 years? 35 years?

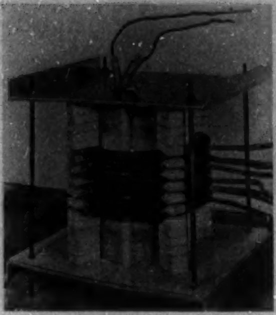


Fig. 5. Dry-type transformer model functional tests

5. What are the effects of radiation of insulating materials used for electronic components? for seals? for capacitors? for motors? What agreement can be reached about nomenclature from radiation? (Such terms as "rad" or "roentgen" units are in use at present.)

6. What are the effects of corona or insulation? What is an accepted definition of corona? Two of the definitions proposed are:

   a. Corona is ionization that is characteristic of partial breakdown in an insulation system.

   b. Corona is an electrical discharge that does not extend from one electrode to another.

7. What are the highest temperatures now considered for insulating materials? Transformers for Air Force requirements are operated at 500 C. What materials can be originated for coverings on wires at such temperature? What capacitor dielectric can be devised? What encapsulating compounds are suitable at 500 C plus the requirement of protection at −65 C? What happens to moisture present at 25 C in cured and uncured systems when temperatures are raised to 500 C in materials, thereby inviting entrapped moisture?

It would seem that, logically, functional evaluation of insulating systems by prescribed test procedures is indicated for electrical equipment. Current questions cannot be answered in a short time. It may take at least 15 to 20 years of work. The designer of equipment essentially must use common sense and assume responsibility. In general, we stand at the crossroads of long-searching investigations.

What Are Basic Problems in Loading of Transformers?

The insulation life of a transformer depends upon the temperature of the hottest-spot in its windings and the length of time this temperature is maintained. Life expectancy of a transformer as presented in the guide for loading oil-immersed distribution and power transformers is conservative.[7] Possibly larger loads for longer duration are permissible without excessive loss of insulation life. But is must be remembered that, by the premises of this loading guide, transformers have established an enviable record of service life. The AIEE working group on the life of materials committee has the task of determining how much transformer insulation life can be expected under accelerated test conditions by following a prescribed test procedure.

In loading oil-filled transformers factors which affect the hottest-spot temperatures are encountered. These factors are:

1. Average temperature rise in windings (copper).
2. Hottest-spot correction for windings.
3. Top oil rise at 100% loading.
4. Ratio of copper loss to core loss.
5. Magnitude of overload.
6. Duration of overload and time constant.
7. Temperature rise at beginning of overload.
8. Ambient temperature at overload.

Omitting the details for obtaining them, it may be said that these factors represent a complex subject. In general, the ASA loading guide[7] on a curve basis gives the per-cent loss of life for hottest-spot temperatures and top oil rises obtained from tests or calculations. Specific transformer constants such as temperature rise, copper loss ratios, and the other items listed are utilized to obtain estimated life. How much these estimates vary from other methods must be determined by the AIEE group. Until these data are available, the user must decide which published data will meet his re-

Fig. 6 (above). Models subjected to temperature aging

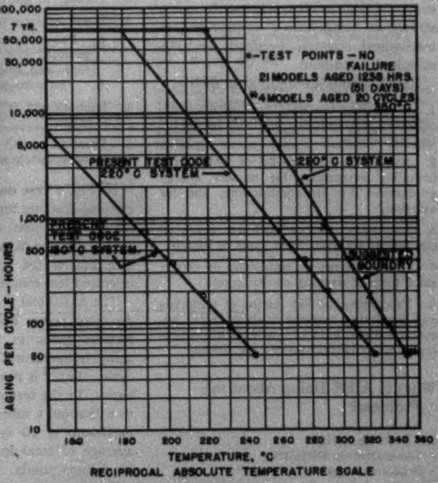


Fig. 7 (right). Test data plot for operating temperature limit; reciprocal absolute temperature scale. Group III, 220 C transformers

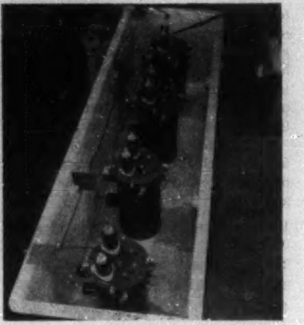


Fig. 8. Accelerated life tests on 15-kva 7,200-120/240-volt transformers

quirements, such as those of Fig. 3, plotted under the conditions assumed.

What Are Basic Considerations in Preparation of Standards for Evaluating Life of Equipment?

Life of electric equipment depends on first cost, reliability required, maintenance, and the importance of weight or size reduction. A life of 35 years may be expected of large power transformers; that of electronic equipment may be a few years, and that of other devices a few hours or minutes.

Experience records play a predominant role in standardization. Laboratory tests cannot duplicate all conditions found in service as temperature, voltage surges, mechanical stresses, overload, and atmospheric conditions.

Recent trends are noticeable, such as:

1. A growing tendency to rely upon accelerated life tests as outlined by AIEE test procedures.
2. Expanding new fields of application including equipment for nuclear energy production, for air conditioning, and for aviation.
3. Increasing use of enclosed types of equipment and controlled gases or liquids.
4. Increasing use of embedded or encapsuled form of winding construction.
5. Procedures for life testing of materials and insulating systems enabling more accurate estimates to be made of life expectancy under certain service conditions.

What is AIEE Doing About the Problem?

1. Basic materials
   - Committee on Dielectrics
     - Subcommittees
       - Gaseous dielectrics
       - Solid dielectrics
       - Thermal evaluation of insulating materials
       - Liquid dielectrics
       - Corona effects
       - Radiation effects
2. Electric equipment
   - Committee on Transformers
     - Proposed Test Procedure for Thermal Evaluation of Ventilated Dry-Type Power and Distribution Transformers. (Published for trial use)
     - Test Procedures for Thermal Evaluation of Oil-Immersed Transformers. Now in process of formulation)
   - Committee on Rotating Machinery
     - Test procedures for motors
3. Co-ordinating Committee no. 4
   - Revising Standard no. 1
   - Preparing Guides for:
     - Test procedure for evaluating thermal stability of insulating materials
     - Test procedure for thermal evaluation of insulation systems for electric equipment

What Are Some Manufacturers Doing?

By active participation in committee work, by building models having certain insulation systems, and by evaluating these models by prescribed test procedures, test data are published.[10-21] For a transformer silicone insulation system, Fig. 5 illustrates an example of a model, and in Fig. 6 the models are shown under test. The test cycle procedure for the insulation system used in the model is:

1. Heat aging, 48 hours at a given temperature.
2. Cooling to ambient temperature.
3. Proof tests; a proof test consists of
   a. Power factor: High to low to ground, low to high to ground, high and low to ground.
   b. Test voltage of 100%.
4. Humidifying at 90% relative humidity for 24 hours.
5. Proof tests with 1/2 hour after removal from humidity chamber.

In Fig. 7 test data are plotted on reciprocal absolute temperature scale graph paper. The user or manufacturer can judge from the data presented the probable life of the system exposed to accelerated test conditions.

Fig. 9. Life tests on 15-kva 7,200-120/240-volt transformers and test equipment

The AIEE test procedure for ventilated dry-type power and distribution transformers gives, in tabulated form, the aging temperatures and aging time per cycle. Fig. 7 is plotted on this per-cycle basis. In the test procedure, models are to withstand a minimum of 5 cycles or a maximum of 10 cycles, from which an average of total hours can be obtained for failure points. On this basis for a 220 C system, the aging program follows:

Hottest-Spot Temperature, C	AIEE Test Procedure, Aging in Air, Total Hours	
	Maximum	Minimum
320	500	250
250	10,000	5,000
220	50,000	25,000
210	100,000	50,000

By testing 21 transformer models, without failure, for a 220 C system by the AIEE test procedure, the author finds the test procedure quite conservative. If failure is to be expected, the proposed boundary is suggested as shown in Fig. 7. For example, four models withstood 20 cycles (48 hours per cycle) at 350 C without failure. This is 10 cycles over the maximum given in the test procedure.

Of recent interest, accelerated life tests for 15 kva, 7,200-to-120/240-volt oil-filled distribution transformers are under way. In Figs. 8 and 9 the test equipment and models under test are illustrated. The procedure for the thermal cycling and voltage tests for one cycle was as follows:

Thermal

1. Circulate winding current until 220 C. Hold temperature until top oil attains 120 C.
2. Cool.
3. Repeat steps 1 and 2 for five loadings.

Voltage

At termination of thermal cycling, steps 1, 2, and 3 apply:

1. Impulse voltage of 95 kv, $1^{1}/_{2}\times 40$ microseconds.
2. Short circuit: 15 times normal current for 2 seconds.
3. Sixty-cycle voltage tests: high voltage, 34 kv one minute; low voltage, 10 kv for one minute.
4. Induced voltage tests, 1.3×rated voltage; 400 cycles, 18 seconds.

This cycle is repeated until complete failure of the insulation system of the transformers. Then, several lower temperatures are selected for thermal and voltage cycle. Results are plotted on graph paper described and estimated file can be determined for these accelerated conditions. By extending the plotted line back to normal permissible hottest-spot temperature a complete picture is presented.

A test procedure for this accelerated aging program is presently under the guidance of the AIEE committee on transformers, delegated to a subgroup on oil-immersed transformers, which reports to the working group on life of materials. This new philosophy of functional testing liquid-filled distribution transformers appeared in 1955.[28]

By combining field experience of transformers, motors, or other electrical devices under overload or extreme environmental conditions, with data obtained by AIEE test procedures briefly outlined, a new tool of great value to the user and manufacturer is made available.

The Future: What Needs to Be Done? What Can Users Do?

Before the industry stands a tremendous amount of work. The job cannot be done overnight. Years of demanding co-operative effort between users and manufacturers are required inaugurating AIEE test procedures, proving that ratings of machines must be on a temperature basis according to the insulation system within them, and establishing temperature ratings under normal conditions with accelerated test conditions. Excessive time is required at normal temperature conditions to establish life relationships. Accelerated tests must be used to expedite better answers than those now available.

Users of electric equipment can be of great assistance by becoming members of various AIEE committees and taking active part in establishing test procedures, and by keeping accurate records of performance of equipment subjected to overloads or to abnormal environmental conditions.

A 3-dimensional relationship of time (duration of load), magnitude of load and life of insulation systems for apparatus is the answer desired. Such a 3-dimensional graph, in terms of per-cent load, will sort out answers to amortization of life of equipment when applications demand 2 years, 3 years, 5 years, 10 years, or 30 years life.

Co-operative papers written by electric power representatives and manufacturers are now appearing in publications.[29,30] By actual load charts, by record of environmental conditions, by a careful check of mortality rates on equipment, and by co-ordinated effort, the aims expressed in this paper can be realized.

References

1. Temperature and Electrical Insulation, **C. P. Steinmetz, B. G. Lamme.** *AIEE Transactions*, vol. 32, pt. I, Feb. 1913, pp. 79–89.

2. General Principles Upon Which Temperature Limits Are Based in the Rating of Electric Machines and Apparatus, *AIEE Standard No. 1*, June 1947.

3. *Ibid.*, Feb. 1954.

4. Encyclopedia of Electrical Insulation, **Gerard de Senarclens.** *Electrical Manufacturing*, New York, N. Y., vol. 58, no. 6, Dec. 1956, pp. 92–99.

5. Overloading of Power Transformers, **V. M. Montsinger, W. M. Dann.** *AIEE Transactions*, (*Electrical Engineering*), vol. 53, Oct. 1934, pp. 1353–55.

6. Effect of Overloads on Transformer Life, **L. C. Nichols.** *Ibid.*, Dec. 1934, pp. 1616–21.

7. Guide for Loading Oil-Immersed Distribution and Power Transformers, *ASA C57.92*, American Standards Association, New York, N. Y., Jan. 1956.

8. Life Expectancy of Oil-Immersed Insulation Structures, **W. A. Sumner, G. M. Stein, A. M. Lockie.** *AIEE Transactions*, vol. 72, pt. III, Oct. 1953, pp. 924–30.

9. Minimum Cost Loading of Power Transformers, **Arthur Klopfenstein.** *AIEE Transactions*, vol. 75, pt. III, Oct. 1956, pp. 979–83.

10. Research Progress in Dielectrics—1956, **A. E. Javitz.** *Electrical Manufacturing*, Nov. 1956, p. 94.

11. Standards for Transformers. *NEMA Publication No. TRI-1954*, National Electrical Manufacturers Association, New York, N. Y., Par. TRI-2-012, 1954.

12. Proposed Test Procedure for Thermal Evaluation of Ventilated Dry-Type Power and Distribution Transformers. *AIEE Publication no. 65*, Nov. 1956.

13. Functional Temperature Endurance Tests on a Silicone Glass Fiber Insulation System for Dry-Type Transformers, **M. L. Manning.** *AIEE Transactions*, vol. 74, pt. III, Apr. 1955, pp. 91–100.

14. A Study of Models for Use in Evaluating Dry-Type Transformer Insulating Systems, **T. R. Walters, A. L. Scheideler.** *Ibid.*, vol. 75, pt. III, Aug. 1956, pp. 520–27.

15. Functional Life Tests of Insulation Systems for Dry-Type Transformers, **Paul Narbut.** *Ibid.*, vol. 74, pt. III, Dec. 1955, pp. 1057–62.

16. Functional Life Evaluation of Group-2 Dry-Type Power Transformers, **H. G. Zambell.** *Ibid.*, Oct. 1955, pp. 956–70.

17. The Functional Evaluation of Insulation for Small Dry-Type Transformers Used in Electronic Equipment, **R. L. Hamilton, H. B. Harms.** *Ibid.*, vol. 73, pt. III-A, June 1954, pp. 605–10.

18. Motor Insulation Life as Measured by Accelerated Tests and Dielectric Fatigue, **C. J. Herman.** *Ibid.*, vol. 72, pt. III, Oct. 1953, pp. 986–91.

19. Functional Evaluation of Motor Insulation Systems, **G. A. Cypher, R. Harrington.** *Ibid.*, vol. 71, pt. III, 1952, pp. 251–53.

20. New Performance Standards for Electrical Insulation of Rotating Machines, **AIEE Committee Report.** *Ibid.*, vol. 73, pt. III-B, Dec. 1954, pp. 1542–48.

21. Insulation Systems for Random-Wound Motors Evaluated by Motorette Tests, **T. J. Gair.** *Ibid.* (Feb. 1955 section), pp. 1702–07.

22. Proposed Test Procedure for Evaluation of the Thermal Stability of Enamelled Wire. *AIEE Publication no. 57*, Oct. 1955.

23. Electric Breakdown of Perfluorocarbon Vapors and Their Mixtures With Nitrogen, **L. J. Berberich, C. N. Works, E. W. Lindsay.** *Ibid.*, vol. 74, pt. I, Nov. 1955, pp. 660–66.

24. Recent Developments in Transformer Insulating Materials, **M. L. Manning.** *Insulation*, Lake Forest, Ill., Oct. 1956.

25. Aging Evaluation of Dry-Type Transformer Insulating Systems, **H. C. Stewart, L. C. Whitman, A. L. Scheideler.** *AIEE Transactions*, vol. 72, pt. III, Apr. 1953, pp. 267–77.

26. Electrical Insulation Deterioration Treated as a Chemical Rate Phenomenon, **Thomas W. Dakin.** *Ibid.*, vol. 67, pt. I, 1948, pp. 113–22.

27. Thermal Aging Properties of Cellulose Insulation Materials, **G. Marmlow.** *Acta Polytechnica* (*Electrical Engineering*), Stockholm, Sweden, vol. 2, no. 1, Issue 31, 1949, 67 pp.

28. Functional Life Expectancy Tests for Liquid-Filled Distribution Transformers, **A. M. Lockie.** *AIEE Transactions*, vol. 74, pt. III, Oct. 1955, pp. 977–85.

29. Economic Loading of Distribution Transformers, **C. F. Mitchell.** *Proceedings*, American Power Conference, Chicago, Ill., Mar. 21, 1956.

30. Pole-Mounted Distribution Transformers of the Future, **A. J. Pansini, M. L. Manning.** Edison Electric Institute, Chicago, Ill., May 1956.

A New Voltage-Divider Circuit for Semiconductor Rectifiers

I. K. DORTORT
MEMBER AIEE

WITH the advent of semiconductor rectifiers into the high-power field, the problem of voltage division across series-connected elements during the inverse-voltage period becomes increasingly important. The various methods used in the past all left something to be desired in economy and effectiveness. The new method described in this paper has been tested and its advantages over the older methods have been proved. At the same time the degree of protection has been improved.

If manufacturing controls were so improved that the voltage division under various operating conditions was held reasonably close during the normal life of the rectifier cells, no special circuitry would be required. Attempts along this line by selective assembly and code markings for replacements have had some success. How long the initial voltage division will be maintained, and the life of rectifiers in this mode of operation, will require statistical analysis of operating experience over a period of years. Certainly the need for stocking rectifiers of several code classifications will prove burdensome in many instances. If cells must be coded for parallel as well as series operation, the number of classifications is greatly multiplied.

An obvious means of voltage division is the use of voltage-dividing resistors connected across each cell. The current in these resistors must be considerably higher than the highest reverse current that will be encountered in the rectifier elements supplied initially and for replacement. A significant drop in over-all efficiency results. To keep the extra power loss at a reasonable level, matching is still required.

An exact division of voltage can be obtained by connecting in series two or more complete rectifiers, including their individual transformers. Instead of separate transformers, individual transformer windings for each rectifier can be built into one large transformer. Both methods have been used in cases of necessity. Both methods are obviously uneconomical as regards cost and space, and also increase losses because of the multiplicity of small transformer windings and leads.

It is therefore a logical step to attempt to use one large transformer winding for the full voltage, with voltage division obtained by small auxiliary transformers or transformer windings connected across each cell. Investigation soon shows that the currents produced by the d-c component of the inverse voltage across each cell must be limited by resistors or blocked by auxiliary rectifiers. Furthermore, the inverse voltage is not sinusoidal, and the matching voltages to be applied to the cells by the auxiliary transformer must generally be synthesized in rather complicated circuits.

These difficulties are overcome as shown in the circuit diagram of Fig. 1(A). Here for simplicity a single-phase full-wave rectifier using a center-tapped transformer and two rectifier cells in each arm is shown. The small auxiliary transformer has a 2-to-1 ratio with its primary winding connected across the full secondary of the rectifier transformer, and its secondary connected between the series junction points on each of the two arms of the rectifier. No d-c voltage appears across either winding and the voltage of the series junction points is rigidly fixed.

Fig. 1(B) shows all the voltages of this circuit plotted with respect to the neutral point of the rectifier transformer winding. The d-c voltage of the output terminal C is shown in heavy line. Note that when phase A is conducting, points A, X, and C are all at the same potential except for the forward drop of the rectifier cells, which is neglected in this paper. Since the potential of point Y is fixed by the auxiliary transformer, halfway between A and B, it coincides with the voltage of the neutral point of the transformer and is so indicated in Fig. 1(B). It is therefore halfway between C and B, and divides the inverse voltage exactly in half. In the next half-cycle, points B, Y, and C are at the same potential and point X is midway between C and A.

The auxiliary transformer can supply no power current to the load because no return path is provided. The increment of reverse current required to bring the inverse voltages of rectifiers $X1$ and $Y1$ to one half the total inverse voltage flows through the secondary of the auxiliary transformer and the common connection between rectifiers $X1$ and $Y1$. The increment of reverse current required to bring the inverse voltage of rectifiers $X2$ and $Y2$ to one half the total inverse voltage flows conductively through the primary and secondary windings of the auxiliary transformer. The current supplied to the auxiliary transformer is therefore, at any instant, the algebraic sum of the incremental currents and the magnetizing current of this transformer.

It is apparent from Figs. 1(A) and 1(B) that under normal operating conditions the inverse voltages of rectifiers $X1$ and $X2$ are fixed at one half the commutating voltage by the secondary of the auxiliary transformer; the inverse voltages of rectifiers $X2$ and $Y2$ are again fixed at one half the commutating voltage by the difference between the two winding voltages of the auxiliary transformer. This system can be expanded to include any number of series-connected cells.

In case any one of the four rectifier cells of Fig. 1(A) fails to block during the inverse-voltage period, the auxiliary transformer is short-circuited even though the

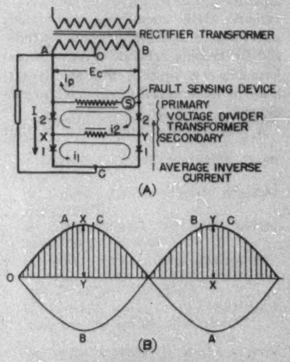


Fig. 1. Single-phase rectifier

A—Circuit diagram of rectifier with two cells in series and voltage-divider transformer
B—Voltage graphs of rectifier

Paper 57-395, recommended by the AIEE Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Southern District Meeting, Jackson, Miss., April 3–5, 1957. Manuscript submitted January 2, 1957; made available for printing February 15, 1957.

I. K. DORTORT is with the I-T-E Circuit Breaker Company, Philadelphia, Pa.

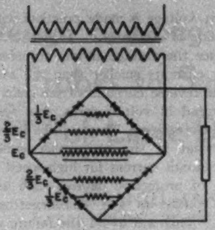


Fig. 2. Single-phase bridge rectifier with three cells per arm and voltage-divider transformer

rectifier transformer does not yet see a fault. The amount of short-circuit current flowing in the auxiliary transformer is determined by the design of this transformer and can be used to actuate a fault-sensing device connected in its primary.

This circuit can be readily expanded to a single-phase bridge rectifier or to any of the well-known multiphase single-way or double-way rectifiers. Fig. 2 shows the application of this voltage-divider circuit to a single-phase bridge rectifier with three cells in each arm.

In Fig. 3 this scheme is applied to a 3-phase single-way rectifier. The double-*Y* connection with the interphase transformer merely requires the duplication of this circuit on the second *Y*. If operation below the transition load is required, the circuit of Fig. 4 can be used. The auxiliary transformers must then be designed for the fundamental and triple-n harmonic voltages.

A 3-phase bridge or double-way rectifier having four rectifier cells in series in each arm is diagrammed in Fig. 5. Eighteen auxiliary secondary transformer windings are required, rated at one quarter, one half, and three quarters of the commutating voltage. The inverse voltage across any cell is fixed by the difference between two consecutively rated windings; i.e., one quarter E_c.

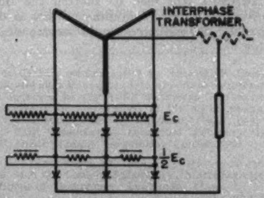


Fig. 3. Three-phase single-way rectifier, two cells per arm and voltage-divider transformers

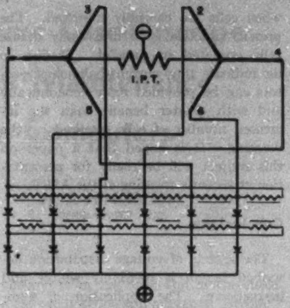


Fig. 4. Double-Y rectifier with interphase transformer

The job can be done with 18 single-phase 2-winding transformers, or a single multiple-winding 3-phase transformer, or other arrangements between these limits.

The voltage division across the remaining good rectifier cells, after one has failed, depends on the impedances of the transformers. At present, it appears that the best control of these impedances can be obtained at the least cost in 2- or possibly 3-winding transformers. For simplicity, Fig. 5 makes use of three single-phase 7-winding transformers with one fault-sensing device in each primary. If more transformers are used, the primaries of all transformers in one phase are connected in parallel to a single fault-sensing device. While Fig. 5 shows the voltage-divider transformers connected delta-delta, it is apparent that Y-Y transformers can be used if there is any advantage in doing so.

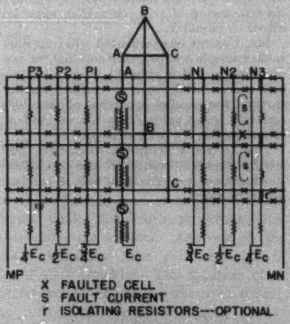


Fig. 5. Three-phase bridge rectifier, four cells in series, two in parallel, with voltage-divider transformers

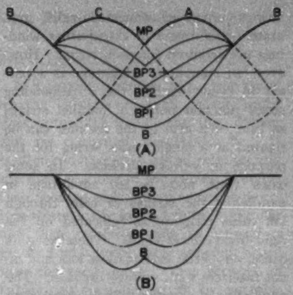


Fig. 6. Potentials and inverse voltages of series junction points of 3-phase bridge rectifier

A—Potentials
B—Inverse voltages

Fig. 6(A) shows the potentials of the three series-junction points of the positive arm of phase *B* with respect to the a-c neutral. Note that the breaks in the curves occur at the time of commutation between phases *A* and *C* on the positive side. Overlap has been neglected in these figures. The inverse voltages with respect to the positive terminal are illustrated in Fig. 6(B).

To demonstrate better the rigidity of the inverse voltages, Figs. 7(A), (B), and (C) show the delta voltages of the junctions maintained at all times by the secondary windings of the auxiliary transformers, with all the deltas shifted to apex *A*, apex *B*, and apex *C* respectively, when phases *A*, *B*, and *C* are conducting.

If the rectifier is made up of parallel

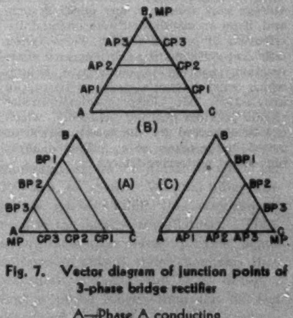


Fig. 7. Vector diagram of junction points of 3-phase bridge rectifier

A—Phase A conducting
B—Phase B conducting
C—Phase C conducting

strings of series-connected cells, separate transformers or transformer windings are required for each string. However, an immediate advantage is apparent in tying together the corresponding series junction points of parallel strings so that each arm of the rectifier can be treated as a single series string for the purpose of voltage division. In some cases there may be an advantage in paralleling the strings of cells through low-value resistors as indicated on the right-hand cells of Fig. 5. These resistors cause very little extra power loss and do not interfere with voltage division. The paralleling of rectifier cells produces very wide unbalances of forward current even when cells are carefully matched. The present practice is to drastically derate cells operated in parallel. Studies so far indicate that current-balancing reactors can be provided more economically and with greater benefit than the increased number of cells required by the derating. It is hoped that a paper on this subject will be ready for presentation at an early meeting of the AIEE.

Conclusions

The system of voltage distribution described herein is relatively simple and inexpensive. The application of semiconductor rectifiers to high-power units of any voltage is made practical without sacrifice of efficiency.

Since the voltage-divider transformers carry currents smaller than the average of the reverse currents of the cells, they are rated in volt-amperes even for the largest rectifiers. They will be increased somewhat in size to provide sufficient short-circuit current for high-speed actuating of protective devices such as bypass contactors and disconnecting devices. This will avoid the failure of all cells in a series string before one bad cell can be detected. With the progressive increase in rating and cost of individual rectifier cells, a considerable decrease in maintenance costs is indicated.

Discussion

J. B. Pitman (Aluminum Company of America, Massena, N.Y.): Mr. Dortort is to be congratulated for his fine paper, a definite contribution to the rectifier art. It is very timely since he has not only developed a practical solution to the problem of equal distribution of inverse voltage, presented to rectifier cells in series, but, from the user's point of view, has brought into sharp focus not only the matter of voltage distribution, which is related to reliability, but also the perplexing matters of "grades" and "matching."

As long as cells must be derated to protect the operation of high-, or large-power rectifiers, which may consist of several thousand cells, the users must pay several penalties. First, cost is high since many expensive cells, which cannot be loaded to capacity, must be provided for protection against the unequal distribution of inverse voltage, and for operation of unmatched cells in parallel. A hazard not present in the operation of mercury-arc rectifiers, against which semiconductor rectifiers must be protected, particularly large installations in critical service, is the occurrence of abnormally high voltage. Occasionally, high voltage may appear on any power system, and, if on an installation with unequal distribution of inverse voltage, across series cells, one cell fails, it might trigger a chain reaction resulting in cumulative failures which could destroy a major portion of an installation. Mr. Dortort's voltage-divider circuit provides positive protection which can be enhanced by the incorporation, as stated in the paper, of co-ordinated ultra-high-speed protective devices.

Another factor, disturbing to present users of conversion equipment, is the matter of grades, resulting in the necessity of matching cells. Unfortunately, as yet, semiconductor rectifier cells do not lend themselves to exact manufacturing controls.

This presents a burden and possibly an intolerable annoyance to the operators of large installations, who must stock many spare units of a multitude of grades, or classifications. The user wants and will insist upon having an installation in which he can, at any time in the life of the installation, replace any cell or a string of cells, or a series-parallel group, without the necessity of accurately matching units.

With reference to the question of matching, the grades are initial grades established by test at the factory. What will be the characteristics of the cell, forward drop, inverse drop, in 3, 5, or 10 years? Will the characteristics remain the same under any reasonable conditions of loading and ambient temperatures? How should spare cells be stored? Will their characteristics remain unchanged over the years?

To gain the widest acceptance for semiconductor rectifiers, certain problems must be solved. If they cannot be solved in the manufacturing of cells, then they must be solved by circuitry.

Mr. Dortort has presented what appears to be a simple and practical solution for one problem, and it is thought that his promised paper on the use of current-limiting reactors for parallel operation of cells will certainly be as valuable to the art and as interesting as the present paper.

This is not a complete discussion, but merely presents some of the questions which are apt to occur to a user of conversion equipment.

I. K. Dortort: Mr. Pitman's kind discussion presents the views of one important user of rectifier equipment. It would have been helpful to hear from other users since the manufacturers need their guidance at this stage of the development of semiconductor rectifiers.

Many questions remain to be answered and these answers will eventually have to come from the users. To enumerate a few of these questions:

1. In various classes of service, is it preferable to shut down a unit the instant one cell fails and prevent further failures, or does the need to continuity of service justify the loss of a string of cells?

2. Will the reduction of inventories of graded cells and maintenance costs justify the greater first cost of the equipment required to accomplish this end, and to what extent?

3. The development of cells for higher voltages and currents will reduce the need for special balancing means and the use of graded cells, but the cost of replacements will go up. To what extent should this program be pushed and how far?

There are problems of unit ratings, standardization, regulation, cooling arrangements, etc., subject to many different solutions. Without some guidance from the industry there will be a tendency to go off in many different directions at the same time which, in itself, might be very useful. However, thorough discussion and expression of users' viewpoints will tend to shorten the time before the kind of equipment which is desired by the industry should become available.

Design of Long Passive Oscilloscope Probes to Simplify Maintenance of Large Digital Computers

W. F. SANTELMANN, JR.
NONMEMBER AIEE

A. J. HINGSTON
NONMEMBER AIEE

THE cathode-ray oscilloscope is a powerful tool for the maintenance of digital computers, but its use on a mobile cart is often awkward and does not permit the oscilloscope to be used to its fullest advantage. If, however, the oscilloscope could be permanently mounted at the maintenance console and supplied with signals from a network of long probe lines running into the computer, operators at the console could view a comprehensive display of computer wave forms and direct maintenance activities accordingly.

Cathode-follower probes driving long, terminated coaxial cables to a central oscilloscope have been used in the past. The cathode followers may be eliminated from such systems by substituting completely passive *R-L-C* networks, which drive unterminated resistance-wire coaxial lines as long as several hundred feet. A 227-foot passive-probe system has been designed for the SAGE *AN/FSQ-7* computer which displays a 16-mμsec (millimicrosecond) rise time, 1% transient distortion, and 30-$\mu\mu$f (micromicrofarad) input capacitance at an attenuation ratio of 100 to 1, when used in combination with a Tektronix *540*-series oscilloscope. A second probe, 55 feet long and of a similar design, is now used in a medium-sized computer, the Memory Test Computer (MTC) at Lincoln Laboratory. This probe has a rise time of 15 mμsec, transient distortion of 2%, input capacitance of 15.5 $\mu\mu$f, and attenuation of 50 to 1.

The resistance-wire center conductor in the coaxial cable absorbs energy reflected from the open end of the cable at the oscilloscope and prevents "ringing" caused by multiple reflections. The severe transient distortions produced by the resistance wire fortunately respond well to equalization by a few *R-L-C* components at the probe. The presence of component tolerances in the equalizer is not critical.

Use of the Maintenance Oscilloscope

Mobile Operation

Maintenance oscilloscopes are normally placed on mobile carts so that they may be moved quickly to a suspected source of trouble in a large computer installation. The method is useful but often requires excessive time to trace a malfunction. Furthermore, wave forms occurring at widely separated points cannot be compared. If a wave form is to be traced for an appreciable distance, the oscilloscope must be moved and warm-up delays occur when its power is interrupted. Most important, the wave-form displays are normally not visible to the operators at the maintenance console who are in the best position for controlling the computer and for diagnosing its faults.

Fixed Operation at the Console

A number of advantages are obtained when signals are transmitted to an oscilloscope in fixed operation at the maintenance console by means of a network of probe lines. The maintenance-console operators can view critical wave forms while diagnostic programs and procedures are applied. Any one of the many incoming lines can be selected for display quickly. The location of a probe can be changed without incurring an additional warm-up period. If a dual-gun oscilloscope is used, direct, simultaneous comparison of wave forms from any two points in the computer can be made.

Long Passive Probe Design

Physical

It is usually desirable to provide equal lengths of cable in each of the probe lines so that the wave forms they deliver will be compared simultaneously. It is often convenient to install permanently the main sections of the cables, bringing their sending ends out to coaxial connectors strategically located in the computer and their receiving ends to connectors grouped on a probe-cable selection board near the oscilloscope. This network of cables can then be driven by portable probes, each carrying several feet of cable to a connector. A minimum of three such probes are required for a dual-gun oscilloscope, two for display wave forms, and one for a sweep trigger wave form. Additional probes will permit wave form switching at the console.

Electrical

The input impedance at the probe tip should be kept as high as possible. The use of an unterminated coaxial cable permits the d-c input resistance to rise to almost any value; 10 megohms is readily achieved. However, the cable capacitance becomes very large in long cables, and only through the use of a very high attenuation ratio is it possible to transform this large capacitance into an acceptably low input loading capacitance. In practice, an attenuation ratio of 50 to 1 for a 100-foot cable or, in general, a ratio numerically half the length in feet, will produce an input capacitance of about 30 $\mu\mu$f. Higher attenuation ratios will produce lower input capacitance, but not in a simple inverse proportion due to the moderating influence of stray capacitance within the probe housing.

Under the assumptions that 30 $\mu\mu$f is the maximum permissible input capacitance and that at least 2 centimeters of deflection are required on an oscilloscope with 50-millivolt-per-centimeter sensitivity, the relatively low sensitivity of these long passive probes requires a minimum signal of about 5 volts per 100 feet of cable. Vacuum-tube computers readily fulfill this requirement even in the largest installations. Transistor computers often provide signal voltages of only 3 volts; however, they are more compact and it is doubtful that a probe longer than the 60 feet permitted under the foregoing assumptions will be required. Consequently, the relatively low sensitivity of long

---

Paper 57-542, recommended by the AIEE Electronic and High Frequency Instruments Committees and approved by the AIEE Technical Operations Department for presentation at the AIEE North Eastern District Meeting, Pittsfield, Mass., May 1–3, 1957. Manuscript submitted February 6, 1957; made available for printing April 5, 1957.

W. F. Santelmann, Jr., and A. J. Hingston are both with the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass.

The research in this paper was supported jointly by the U. S. Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology.

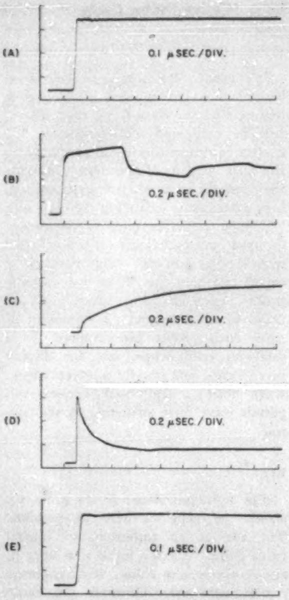


Fig. 1. Basic long passive probe wave forms

Step input (A) is transmitted by 227 feet of unterminated cable as in (B); by resistance-wire cable in waveform (C). The step input is reshaped by the attenuator-equalizer of Fig. 2 in waveform (D) and applied to the resistance-wire cable for output (E)

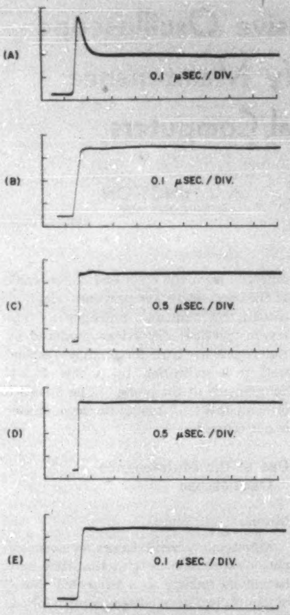


Fig. 3. Operation of attenuator-equalizer

Waveform (A) is the resistance-wire cable response to a step input when driven by a conventional attenuator (R1, R4, C1). The large overshoot is controlled in (B) and (C) by R5 and C2. The bulge is reduced by R6, R7, C3, C4, and L1 in (D) and (E)

passive probes appears to be no serious disadvantage.

Design Specifications for the AN/FSQ-7 Probe System

In order to justify changing an existing cathode-follower probe system to a passive design, it was necessary to design a probe 227 feet long which would provide satisfactory oscilloscope deflection from input signals of 20 to 40 volts with an input capacitance not appreciably more than 30 $\mu\mu$f and an input resistance greater than 1.0 megohm. Also, the over-all rise time, including the probe system and its oscilloscope, was required to be less than 20 mμsec, the transient distortion less than 3% peak, and the signal-to-noise ratio better than 20 decibels. "Noise" here refers to stray computer-pulse and power-frequency voltages which are coupled into the long lines, rather than to the random noise of information theory. In addition, the probe was to be direct-coupled for transmission of flip-flop level information. And finally, it was to be feasible to place coaxial connectors in at least two places in the cable and to coil unused lengths of cable without significant transient distortions arising from either act.

Development

The development of the long passive probe was stimulated by the performance of the Tektronix *P410* probe, a 42-inch unit which utilizes resistance wire as the center conductor of the cable to damp cable ringing and produces very faithful wave-form transmission. The first attempt at extending the length of resistance-wire passive probes was to lengths of 5 to 8 feet, but gradually the lengths were extended to the full 227 feet. It was fortunate that Federal *K-109A* cable is manufactured with a center wire suspended by crimping in a relatively large bore (about 0.080-inch diam) because it was a simple matter to solder a resistance wire to one end of the phosphor-bronze center wire which was pulled out of the cable. No breakage or jamming ever occurred, even in the 227-foot lengths. It was therefore possible to study the performance of a number of cable resistances on a "do-it-yourself" basis.

The cables made in this way are completely satisfactory except for an electrostatic voltage generation which occurs with cable movement and may reach 0.2 volt. If the cable is motionless, as it would be in a cable trough, there is no problem. A special resistance-wire cable, type *CE-330*, developed by Federal Telephone and Radio Corporation for this probe system, exhibits almost none of this voltage generation.

It was discovered that cable resistances ranging between 350 and 700 ohms total were required to damp out cable reflections with minimum degradation of the transient rise. Thus, the resistances per foot of various lengths of cable must differ in order that the total resistance will lie between 350 and 700 ohms, and it is not practical to use only one type of resistance-wire coaxial cable for all probe designs.

Fig. 1 illustrates the waveforms associated with the long passive probe. Waveform 1(A) is the output of a

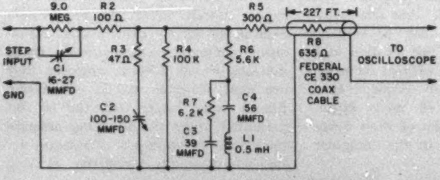


Fig. 2. The 227-foot passive probe circuit

Fig. 4. The 227-foot passive probe gun

mercury-wetted contact relay used to generate a step with a rise time in an order of a few mμsec, as seen on the oscilloscope with a 12-mμsec rise time. The observed rise time is almost exactly 12 mμsec. Waveform 1(B) is the result of transmitting the step waveform along a 227-foot length of Federal *K-109A* with no termination. The severe reflections which occur are damped out by the addition of resistance wire in the coaxial cable in waveform 1(C). Although cable reflections are no longer a problem, the knee in the rise of waveform 1(C), caused by skin-effect impedance on the resistance wire, is severe. The knee distortion increases with the length of the passive cable. Cable lengths up to about 5 feet, when compensated by a single small series inductance at the oscilloscope end, produce no noticeable knee on a 12-mμsec-rise-time oscilloscope.

In order to transmit the step waveform faithfully, it is necessary to insert between the step waveform and the cable a means of reshaping the step as in wave form 1(D). The combined transient responses of waveforms 1(C) and 1(D) produce a reasonably accurate output wave shape as in Fig. 1(E). This reshaping device is called the attenuator-equalizer because it also provides the probe attenuation ratio. The circuit devised for the 227-foot probe is shown in Fig. 2.

The circuit of Fig. 2 is completely passive and may be constructed from extremely small components. Power dissipation is very low, a small fraction of a milliwatt at most, so the physical size of the components is determined primarily by the voltage ratings of resistors and capacitors. The function of each of the elements of the attenuator-equalizer is explained by reference to Fig. 3. Waveform 3(A) illustrates the output of a resistance-wire cable when driven by a standard attenuator comprised of *R1*, *R4*, and *C1* of Fig. 2. The large overshoot is controlled in waveform 3(B) by the addition of *R5* and *C2*. Capacitor *C2* is made adjustable so that the height of the leading edge may be controlled.

Waveform 3(C) is identical to 3(B), except that it is taken at a sweep speed five times slower to accentuate the bulge which occurs during the first 0.5 μsec after the rise. The bulge is reduced considerably by the addition of the network comprised of *R6*, *R7*, *C3*, *C4*, and *L1*, as illustrated by the slow-sweep waveform of Fig. 3(D). The total influence of these five components is only a few per cent. It is often feasible to omit the entire network if transient distortions of up to 5% are tolerable. Waveform 3(E) is the final circuit response at 0.1 μsec per division including the two additional components *R2* and *R3*. The function of *R3* is to decrease the rise time while *R2* damps ringing which may occur in the input leads when viewing extremely fast wave fronts.

The tolerances on the fixed components are not unduly severe. Resistors *R1* and *R4* must be 1% units for accuracy of d-c attenuation. Otherwise, the fixed components require no more than the customary 5% tolerance, and some will perform consistently with 10% tolerance. A given attenuator-equalizer circuit is capable of performing well with cable resistances within 5% of its nominal design-center value. With suitable variations in the attenuator-equalizer, cable resistances between 350 and 700 ohms per cable length may be used, but the region between 400 and 500 ohms appears to offer the greatest performance potential for a minimum circuit complexity. Although a 400- to 500-ohm cable does not in itself provide complete absorption of the first reflection, an additional absorption of this reflection in the attenuator-equalizer reduces its amplitude below the resolution of the oscilloscope.

Probe Performance

The 227-foot passive probe attenuator-equalizer was built into the gun illustrated in Fig. 4 for use in the prototype air defense computer at Lincoln Laboratory. The probe gun carries 27 feet of coaxial cable to a connector which plugs into a permanently mounted jack on the computer frame. A network of similar cables converges to a "probe-cable selection board" mounted next to an oscilloscope at the maintenance console as illustrated in Fig. 5. Three

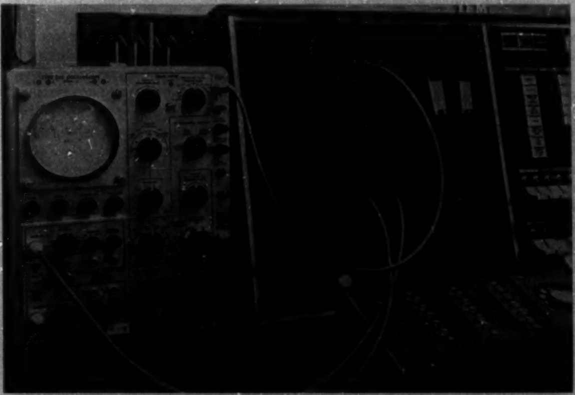
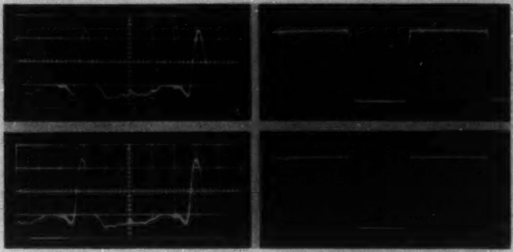


Fig. 5. Probe-cable selection board



(A) 0.1 μsec/cm (B) 1 μsec/cm

Fig. 6. Probe response comparison on computer waveforms

A—Tektronix P410 (42-inch probe)
B—The 227-foot passive probe

short cables from the oscilloscope permit the selection of any three incoming waveforms, two of which are viewed while one is used to trigger the sweep.

Typical computer trigger pulses and flip-flop waveforms are used in Fig. 6 to compare the performance of the 227-foot probe with that of the standard 42-inch Tektronix *P410* probe. The *P410* probe waveforms on the top differ only slightly from the 227-foot probe waveforms on the bottom. The transient distortions of the long probe, while certainly not zero, are insignificant when the long probe is used in maintenance work.

Reflections caused by the insertion of coaxial connectors in the cable are also damped by the resistance wire and cause no more than 1% transient distortion. Coiling the modified *K-109A* cable often caused a 5–10% amplitude ringing which could be removed by separating turns by about one inch. The Federal *CE-330* cable has very little between-turn coupling and requires no special spacing.

Table I tabulates the measured performance of the 227-foot probe as well as that of a 55-foot passive probe developed for the MTC. All measurements, except input loading, were taken on a Tektronix *540*-series oscilloscope. The rise time of the longer probe can be reduced to about 14 mμsec by the omission of *R2* which is used to damp ringing in the input leads. This modification requires that the input leads be no longer than about 2 inches if the fast rise time is to be obtained without ringing in the leads.

Accurate measurement of the transient distortion required near-perfect adjustment of the oscilloscope distributed amplifier. It seems entirely feasible, if a particular oscilloscope is to be used only with long probes of closely similar performance, to align the oscilloscope for best wave shape when a fast step waveform is applied to the probe rather than the oscilloscope itself. In this way, the distributed amplifier delay line may be made to balance out transient distortions of a few per cent during the first 0.4 μsec after a fast rise.

The signal-to-noise ratio measurements are estimates made on the basis of computer waveforms of 40-volt peak-to-peak amplitude compared with the stray-pulse and power-frequency signals received when the probe input is grounded at the remote point and the computer is in full operation. Since the probe gun must be grounded at the waveform source and the oscilloscope must also be grounded, it is often found that power-frequency voltages on the computer ground bus are delivered by the probe system to the oscilloscope without appreciable attenuation. In such cases, a capacitor of about 0.001 μf, in series with the probe gun ground lead, will block the flow of power-frequency signals without disturbing computer waveforms. The direct-coupling feature is still maintained by virtue of the common grounding of the waveform source and oscilloscope through the computer ground bus. The pulse "noise" which remains is usually locked in phase with the waveform being displayed and therefore will be observed as a very slight change in wave shape rather than as an interfering pulse train.

Table I. Long Passive Probe Performance

Probe	Attenuation Ratio	Rise Time, mμsec	Transient Distortion, %	Input Loading		Typical Signal-to-Noise Ratio, Decibels
				Megohms	$\mu\mu$f	
227-foot	100 to 1	16	±1	9.0	30	38
55-foot	50 to 1	14	+2, −1	9.0	15.5	>50

The relatively low input capacitance of the 55-foot probe is partially a result of a simpler, 11-component circuit, constructed compactly inside a Tektronix *P410* shell. The 227-foot probe gun contains a total of 15 separate components.

Conclusions

Cathode-follower probes can be replaced with entirely passive attenuator probes constructed using resistance-wire coaxial cable. Desirable characteristics are maintained with cables up to several hundred feet long. The usefulness of long passive probes as a maintenance tool for electron-tube or transistor digital computers is not diminished by their relatively low sensitivity and high input loading. The attenuator-equalizer circuit is simple and reliable but increases in complexity with the length of the probe cable. The cable damping resistance for any length of cable should be between 400 and 500 ohms, although resistances from 350 to 700 ohms can also be made to perform well. A 55-foot probe system has been in operation in the MTC for nearly a year, and a 227-foot system has been installed in the prototype SAGE air defense computer at Lincoln Laboratory for several months.

A Resistance-Temperature Detector for Nuclear Reactor Service

F. R. SIAS
ASSOCIATE MEMBER AIEE

SCIENTIFIC ADVANCES are continually making new demands on old arts. This is particularly true of the field of atomic energy, where instrumentation is finding itself hard pressed to keep in step. The inexhaustible source of power is at hand, but controlling it so that it can efficiently do its work is taxing our ingenuity.

In a "heat" power plant, temperature is a major parameter. The higher the temperature of the power source, the greater will be its efficiency. And if a liquid is used as a working fluid, the pressure will increase with temperature and approach the critical pressure. The accuracy of temperature measurement determines whether work can be done efficiently, close to the upper boundaries of mechanical strength and close to critical pressures, or whether large safety margins must be left.

Various systems are available for the measurement of temperature. For application in a nuclear reactor, fluid-filled systems are excluded, because in these the connecting piping would make a passage for the escape of radioactive materials. Thermocouples can be made to give reasonably good accuracy. But the systems are complicated by cold junction compensation devices, and the signal power available is small. The electrical resistance thermometer has long been the standard for precision thermometry. Much more power can be used in the measurement circuits than in a thermocouple system, and no cold junction compensation is required. It was therefore selected for this critical measurement task.

Specifications

If water is used as the heat exchange medium in a reactor, the critical pressure and temperature is 3,184 psi (pounds per square inch) and 705 F (degrees Fahrenheit). These were used as preliminary design goals. Additional requirements include a flow rate of 45 feet per second, and a response time of 1.0 second (63%). Devices used near a reactor are also subject to a radiation field, the intensity of which depends on the location of the device. For this application a fast neutron flux of 1 × 10^5 n/cm^2 sec (neutrons per square centimeter per second), thermal of 5 × 10^8 n/cm^2 sec, and gamma, 200 r/hr (roentgens per hour) was selected as an upper limit. Desired life is 40,000 hours.

The ambient air temperature is 70 C (degrees centigrade). Navy hi-shock and vibration tests are required. Also, equipment attached to submarine piping must be strong enough to be used as a step for maintenance people in addition to its normal function because of the extremely crowded conditions, especially in the reactor compartment.

Design

Sensing Element: Radiation Effects

There are always conflicting requirements for an RTD (resistance temperature detector) sensing element. Fast response demands low thermal capacity and high thermal conductivity. Low self-heat demands large areas. Strength calls for thick protecting walls. Good electrical stability requires that electrical insulation be good, usually with low thermal conductivity. Wire of soft material such as platinum must be mounted so as to be strain-free with attendant poor thermal contact. These qualities must be balanced in a good RTD to best fulfill the design goals.

First, of course, are its electrical characteristics: its resistance-temperature curve. This must be precisely established, and it must be stable over a long period. Studies which have been previously reported[1] show the advantages to be obtained from the use of tungsten as the resistance material. Stabilities obtained are as great as those of platinum, and in addition, precise control of temperature coefficient and resistance is more easily obtained.

In addition to stability under operating temperature conditions, the RTD must be stable under a radiation field for a number of years. Nine tungsten units were exposed to radiation in the instrument tunnel at the Brookhaven National Laboratory for a period of 40 hours, to a radiation field which gave an integrated flux the equivalent of 40,000 hours of operation as follows:

Neutrons	0.025–0.4 ev*	2×10^9	n/cm^2/sec
	1–10 ev	7×10^7	
	10–100 ev	7×10^7	
	100 ev–1.5 mev†	2.7×10^8	
	1.5–2.5 mev	6.6×10^6	
	2.5–6.0 mev	3×10^5	
Gamma		5,600 r/hr	

* Electron volts. † Megelectron volts.

Twelve additional units were maintained at about the same ambient temperature to act as controls. At the end of the test period, measurements were made on all 21 units with these results given in the table below. There were no significant changes due to radiation.

Fast response requires that the sensing resistance be in as intimate contact as possible with the medium whose temperature is to be measured. In addition, it must withstand very high pressures. The construction chosen is shown by Figs. 1 and 2. The winding itself is wound on three notched mica cards, as shown, and nichrome leads are attached. Nichrome is used because it has a very

	Temperature, C	Change	
		Average	Maximum
9 irradiated RTD's	0	+0.006	−0.05
	100	−0.002	−0.03
12 controls	0	+0.001	−0.06
Maximum uncertainty of data			±0.04

Paper 57-901, recommended by the AIEE Nucleonic Radiation Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the 2nd Nuclear Engineering and Science Conference, Philadelphia, Pa., March 11–14, 1957. Manuscript submitted March 19, 1957; made available for printing May 22, 1957.

F. R. Sias is with the General Electric Company, West Lynn, Mass.

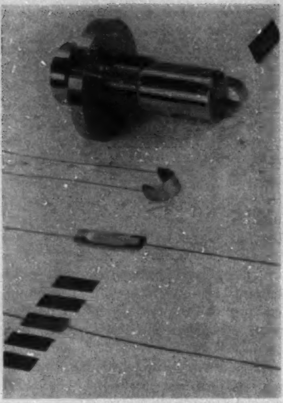


Fig. 1. Construction of RTD, showing winding wound on notched mica cards and attached nichrome leads

low thermal electromotive force against tungsten over a very wide range. Two additional mica strips are placed on either side of the winding to form a sandwich. Several thin (0.0025-inch) mica strips are used instead of single thicker ones, to facilitate the bending operation which follows.

The sandwich is then slipped into a flattened stainless-steel tube and the whole bent in a die to the U-shape shown. This operation stresses the outside wall of the U beyond the elastic limit and leaves the walls of the tube pressing on the mica sandwich, with considerable pressure, for best thermal contact.

The sensing element, Fig. 1, is next inserted into the machined groove in the nose of the mounting block, Fig. 2, and brazed with a gold-nickel alloy. This method of mounting leaves the tube supported both on the inside and outside. When pressure is applied to the sensing element, it does not put the gold solder in tension. The remaining space in the sensing element, around the leads, is filled with alumina powder, packed as tightly as possible. Thus external pressure cannot distort the protecting tube, but instead tends only to increase the pressure on the sensing element and improve its response characteristics.

If the sensing winding were made of platinum or other soft metal, the pressure on it would seriously strain and distort the wire, thus changing its resistance characteristic. Even though the adjustment were perfect before forming, and stable after forming, it would have a significantly different value, and the amount of change during forming could not easily be predicted. When tungsten is used, its high strength allows little change in resistance during the forming process. In addition, since the resistance adjustment is made chemically on the completed element, as explained in the previous article,[1] any change that does occur is corrected. Tests have been made on the completed RTD which show that the temporary change in resistance due to external pressure is negligible.

Tungsten has a further valuable advantage. When an RTD is used for indicating or controlling temperature, it is often used in a bridge circuit which is balanced at one particular temperature, and at any other temperature is unbalanced. The voltage ouput of this bridge, for a uniform change in resistance of the RTD, is nonlinear. It is desirable for many applications to have an output voltage that is linear with temperature. This is possible with a tungsten RTD, for the bridge and RDT curvatures are opposite, and compensate each other when the correct values of bridge resistors are chosen. The linearity can be made better than 1 F over a 700 F range, and to better than 1/4 F over most of the range.

Mounting Block and Welding Boss

As previously mentioned, the sensing element is gold brazed into the mounting block. This block is machined for two sensing elements so that two separate temperature measurements or a spare element can be provided with a single penetration of the high-pressure piping; see Figs. 2 and 3. This penetration is made by welding a boss, part no. 3, shown in Fig. 3, to the pipe. The boss

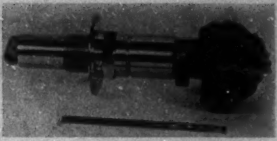


Fig. 2. Construction of RTD, showing sensing element

is threaded internally oversize, and a Helicoil insert is installed. The insert is a helical coil of hard-drawn diamond-shaped wire screwed into the thread to serve as the inner thread surface of the nut. The RTD screws into this thread. It reduces the possibility that thread galling will prevent the removal of the RTD. The thread is used to provide the strength to withstand the internal pressure in the pipe. To prevent leakage, a seal weld, part no. 10, is made at the flange as shown.

The mounting block also contains passages for the leads. Ceramic high-temperature hermetic seals, part no. 9, are silver-brazed directly to this part. An evacuating tube for each sensing element allows powdered alumina to be introduced and packed as much as possible to prevent movement of the leads, due to vibration and shock. Helium is then added at a pressure slightly higher than atmospheric pressure. The sleeve, part no. 4, is then installed and welded (no. 8) in place.

Head

The head must be waterproof. It is also desirable to allow the conduit fitting to be located in any direction. Therefore the head is mounted on the sleeve with an O-ring seal and clamp to secure it

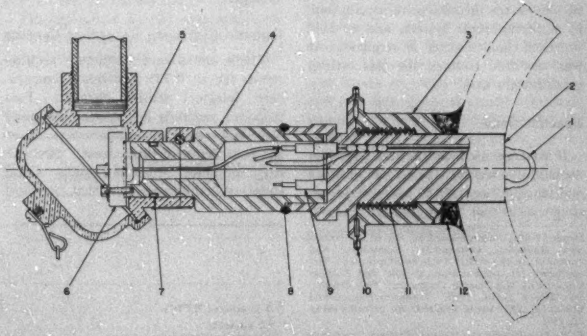


Fig. 3. Diagram of RTD showing various parts

in the desired position. The terminal board is mounted directly on the sleeve to allow rotation of the head without twisting the leads. Before assembly of the terminal board, the cavity in the sleeve is filled as completely as possible with coarse alumina to prevent failure of the leads due to shock and vibration.

Testing

RESPONSE TIME

Response time tests were made by inserting the RTD in a pipe in which water at about room temperature could be maintained at flows up to 50 feet per second. Water was first drained from the pipe and the temperature of the RTD raised to about 100 C. A quick opening valve allowed a full flow of water to enter the pipe, thus making a step change in temperature. The RTD was connected to a Mueller bridge balanced for a 63% change in temperature. It was excited with 60-cycle alternating current and the output was applied to a recorder. The response time was taken as the time between the beginning of the change in bridge balance to the balance point. At a flow of 30 feet per second, the response was 0.4 second, and at 5 feet per second it was 0.6 second.

SELF-HEAT

This test was made in the same apparatus as was used for response time. Flow was set at a desired rate and the RTD resistance was measured with negligible power applied to the bridge. The power was then raised and the change in resistance was converted to self-heat temperature rise. At 15-feet-per-second flow and one watt dissipated in the RTD, the rise was 1.7 F. At 0.1 watt, the rise was 0.2 F.

PRESSURE TESTS

The various components of the RTD have been pressure tested separately. This was done on a dead-weight tester, failure being indicated when an actual leak occurred, or where it became impossible to increase the pressure further due to the expansion of the system.

The sensing element itself was tested to 5,500 pounds/in^2 without failure. The only visible effect was a slight re-forming of the base of element where it was brazed into the mounting block. No significant changes in resistance due to pressure were noted. Insulation resistance remained high.

Although the hermetic seals are not normally required to withstand high pressure, in the event of a failure of the sensing element, they would be subjected to full working pressure. A number of seals were tested up to 10,000 psi, the minimum failure pressure recorded being 7,000 psi.

An special test model of the mounting block and boss was made to test the effectiveness of the Helicoil thread and seal weld design. The seal weld showed no sign of failure up to 9,000 psi. The thread began to give at about 8,500 psi.

Conclusions

An RTD has been designed which will more than fulfill the requirements for operation in an atomic submarine. It has very fast response and high power dissipation ability under flow conditions. It can be adjusted to high precision and will maintain this adjustment for long periods in a radiation field.

Although designed for operation at less than 3,000 psi on the sensing element, it has withstood testing well over 5,000 psi. It is sufficiently rugged to withstand extreme shock and vibration.

Reference

1. A TUNGSTEN RESISTANCE THERMOMETER, F. R. Sias, J. R. MacIntyre, A. Hansen, Jr. *AIEE Transactions*, vol. 73, pt. I, Mar. 1954, pp. 66-69

Safety Circuit Development at Brookhaven National Laboratory

J. E. BINNS
ASSOCIATE MEMBER AIEE

W. LONES
ASSOCIATE MEMBER AIEE

D. G. PITCHER
NONMEMBER AIEE

M. MÉLICE
NONMEMBER AIEE

FROM the time of the earliest nuclear reactors, it has been considered of primary importance that they be protected against an uncontrolled rise in temperature, whether this rise be due to a corresponding uncontrolled rise in neutron level or to a failure of the cooling system. While it is claimed that some types of reactor are intrinsically safe from this kind of accident, it is likely that until a solid background of experience has been acquired, all nuclear reactors will continue to be protected by devices which "scram" them when such devices foresee a dangerously high temperature. But temperature-sensing devices are too slow, as well as too insensitive at low levels to provide protection during startup. Therefore neutron-sensing devices are used. The present paper will be limited to discussing that part of the safety system which senses neutrons. Similar principles are applicable as well to other portions of the safety system, such as those which sense temperature, coolant flow, etc.

Safety systems generally are subject to two requirements:

1. The frequency N of false scrams must be small.
2. The frequency n of failure of the reactor to scram in spite of a genuine scram signal must be extremely small.

It will be recognized that neither N nor n can in practice ever be made actually zero. Rather, one is faced with the task of using the various means at his command to reduce each of these numbers to an acceptable value. One must first decide in each case what shall be the acceptable value. Then after practical solutions have been studied, the law of diminishing returns may dictate a revision of these estimates.

One must then examine the various factors which may be used to minimize N and n. Typical of such factors are the use of coincidence trip systems, the use of automatic testing, and the elimination of means for disabling part of the system. In studying the desirability of using any such factor, questions such as those on the following page should be asked:

Paper 57-398, recommended by the AIEE Nucleonic Radiation Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the 2nd Nuclear Engineering and Science Conference, Philadelphia, Pa., March 11-14, 1957. Manuscript submitted January 25, 1957; made available for printing May 22, 1957.

J. E. BINNS, W. LONES, and D. G. PITCHER are with Brookhaven National Laboratory, Upton, N. Y., and M. MÉLICE is with the Nuclear Development Corporation, of America, White Plains, N. Y.

This work was done under the auspices of the United States Atomic Energy Commission.

Table I. Evaluation of Design and Operating Factors

	Factor	Effect on "N"	Effect on "n"	Effect net
1.	Change an unsafe fault to a safe fault	−	+	+
2.	Add a channel (largely fail-safe)	−	+	?
3.	Add 1 to the number of tripped channels needed to scram	+	−	?
4.	Item 2 plus item 3	+	0	+
5.	Test each channel with simulated signals to detect unsafe faults	−	+	+
6.	Increase the frequency of such tests	−	+	+
7.	Decrease duration of such tests	+	0	+
8.	Make such tests automatic	+	+	+
9.	Automatically disable channel during test and maintenance	+	−	−
10.	Manually disable channel during test and maintenance	?	−	−
11.	Make automatic the interpretation of the results of the test	0	?	?
12.	Make automatic the action taken on the results of the tests	−	?	?
13.	Provide an interlock which disables testing during the time any channel is tripped	+	−	+
14.	Provide a switch to disable testing for a limited time	+	−	+
15.	Provide operator with a meter in each channel to show proximity to trip point	+	0	+
16.	Provide convenient spares for major units of system	+	+	+
17.	Make these spares of plug-in type to minimize time channel is tripped due to safe fault or inoperative due to unsafe fault	+	+	+
18.	Design for minimum routine maintenance (exclusive of operability tests)	+	+	+
19.	Design for simplest operating procedure	+	+	+
20.	Install in full view of operator all equipment which gives information about safety system	+	+	+
21.	Keep operator's attention on safety system by making it an active rather than a dormant part of the equipment, thus insuring adequate monitoring	0	+	+
22.	Set up supply and shop procedures to insure that serviceable spare units be always available	+	+	+
23.	Operate reactor so as to have fewer situations calling for a scram	0	+	+

1. Is the effect on N good, bad, or unpredictable?

2. Is the effect on n good, bad, or unpredictable?

3. If the effects tend to cancel each other, is the net effect good, bad, or unpredictable?

4. What is the point of diminishing returns for this factor? (This will change as other factors are applied.)

One will not use any factor unless its effect on either N or n can be predicted to be good, and unless its net effect can be predicted to be good. Nor will one use any factor unless it produces a significant improvement.

In deciding whether the net effect is good or bad it is obviously not the change in $(N+n)$ which has to be considered, since n will be many orders of magnitude smaller than N and represents a different kind of event. Nor can one consider the sum of the percentage changes:

$$\frac{\Delta N}{N}+\frac{\Delta n}{n}$$

since the two percentage changes have entirely different significances. A reduction of $\Delta N/N$ in the number of false scrams means a dollar saving of the same percentage. But a reduction $\Delta n/n$ in the number of "catastrophes" is of less tangible significance, being of negligible significance if the probability of such catastrophe is already negligible. It should therefore be emphasized that a proper evaluation of the various factors depends more on common-sense estimates based on operating experience than on a mathematical formulation.

In Table I are listed most of the factors which have come under consideration in the design of the present system, together with estimates of their effects. In this table, the symbols have the following meanings:

+ = definitely good
− = definitely bad
0 = no significant effect
? = effect unknown

All the factors receiving a plus in the "net" column have been used in the safety system which will now be described, and which we intend to make a part of the new Brookhaven Medical Research Reactor as well as of the old Brookhaven graphite research reactor.

Description of System

Fig. 1 shows a neutron-sensing system consisting of three channels which are independent as far as, but not beyond, the actual scramming device. In each channel there are two signals: one, proportional to reactor power, which comes from an uncompensated ion chamber, and the other, proportional to inverse period, which comes by way of a period amplifier from a gamma-compensated pair of ion chambers. The actual trip signal is the sum of these two signals.

The result of this combination of signals is a lengthening of the trip period as the power increases. The condition for a trip is

$$\frac{T'}{T}+\frac{P}{P'}>1$$

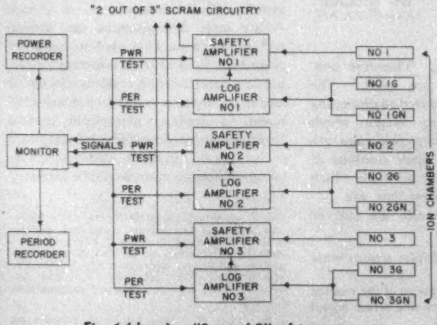

"2 OUT OF 3" SCRAM CIRCUITRY
POWER RECORDER
MONITOR
PERIOD RECORDER
SIGNALS
PWR TEST
PER TEST
PWR TEST
PER TEST
PWR TEST
PER TEST
SAFETY AMPLIFIER NO 1
LOG AMPLIFIER NO 1
SAFETY AMPLIFIER NO 2
LOG AMPLIFIER NO 2
SAFETY AMPLIFIER NO 3
LOG AMPLIFIER NO 3
NO 1
NO 1G
NO 1GN
NO 2
NO 2G
NO 2GN
NO 3
NO 3G
NO 3GN
ION CHAMBERS

Fig. 1 (above). "2-out-of-3" safety system

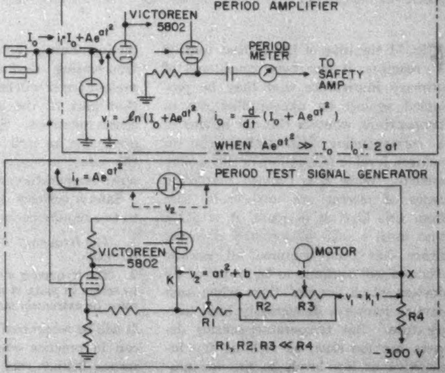

PERIOD AMPLIFIER
VICTOREEN 5802
$I_o \rightarrow i_1 = I_o + Ae^{at^2}$
PERIOD METER
TO SAFETY AMP
$v_1 = \ln(I_o + Ae^{at^2})$
$i_o = \frac{d}{dt}(I_o + Ae^{at^2})$
WHEN $Ae^{at^2} \gg I_o$ $i_o = 2at$
PERIOD TEST SIGNAL GENERATOR
$i_1 = Ae^{at^2}$
v_2
VICTOREEN 5802
I
K
$v_2 = at^2 + b$
MOTOR
X
$v_1 = k_1 t$
R1
R2
R3
R4
R1, R2, R3 ≪ R4
− 300 V

Fig. 2 (right). Period amplifier test signal generator

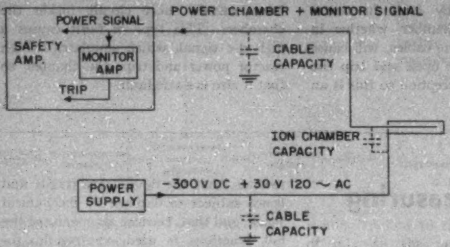


Fig. 3. Ion chamber and cable monitor

where

T' = trip period at $P = 0$
P' = trip power at $T = \infty$

The combined trip signal is indicated on a meter which is calibrated from 0 to 100% of trip point. The advantages of combining signals are the following:

1. Excursions of temperature due to increasing power can be made less dangerous, since the higher the power the sooner will the reactor be scrammed by a shortening period.
2. The operator can see on a single meter at all times how close the channel is to the trip point.
3. The operator can start up and level off simply by holding this meter at a constant reading.
4. The equipment is simplified by the fact that only one trip circuit is needed for both signals.

When this trip signal reaches the set level in any one of the three channels, the plate current in the output tube drops and de-energizes the scram device, as, for instance, a relay, and actuates the scram annunciator for that channel. The scram devices of the three channels are so connected that if any two of them trip, the reactor is scrammed.

The rest of Fig. 1 shows how the system is monitored. Monitoring is not fully automatic, but depends on the operator to observe the charts of the monitoring recorders. The safety monitor keeps a running check of the safety system and automatically performs the following operations:

1. Sequentially switches the period and power recorders to each channel every 20 minutes. This permits the operator to cross-check the power readings of all channels.
2. Sequentially, 5 minutes after switching the recorders, connects a power signal, and 10 minutes later a period signal, to the input of the power and period amplifiers, thus checking the system once every hour.
3. Five seconds before the test signal is applied, disables the scram annunciator for the channel under test. This is done to avoid frequent, meaningless alarms.
4. Marks recognizable pips on the recorder charts, thus allowing the reactor operator to check the system trip calibration at a glance. This is accomplished by de-energizing the power and period recorder balancing motors for set intervals upon a test-induced scram. Specifically, although the trip calibrations and the power amplifier circuits of the safety amplifier are checked exactly, the ion chambers and the calibration of the period amplifiers are not checked. The operability of these amplifiers is checked to the extent that a signal impressed at the input results in a tripping period.
5. Stops the test sequence upon an operational or malfunctional trip in any channel. This prevents a reactor scram which would result from the test.

In addition, provision is made for the manual calibration of any power or period trip circuit, in order that the operator may check the result of an adjustment. In this case the test signal simulating an inverse period is applied at the output, rather than at the input, of the period amplifier.

A switch is provided to enable the operator to disable the automatic testing for a limited period of time, such as during startup, when the testing may be an undesirable distraction.

For reaching the scram point and producing a reasonably clean trace on the charts, linearly rising signals at the two inputs of the safety amplifier are desirable. The test signal at the power input is in the form of a ramp function simulating a linearly rising ion chamber current; however, to get a linearly rising signal at the period input of the safety amplifier, a current of the form

$$i_t = Ae^{at^2}$$

has to flow into the input of the period amplifier. In order to obtain this form of current, a special signal generator is provided. Fig. 2 shows a motorized potentiometer R_3, driven at a constant speed in the output circuit of the period test amplifier, and generating a voltage

$$v_1 = k_1 t$$

This potentiometer, in conjunction with resistors R_1, R_2, and R_4, forms a squaring circuit R_1, R_2, $R_3 >> R_4$, so that the voltage drop from point K to point X is

$$v_2 = at^2 + b$$

This same voltage is also impressed across the test diode, which, being operated in its logarithmic range, results in a current flow through the diode of the form

$$i_t = Ae^{at^2}$$

This is made possible by the inclusion of a feedback amplifier consisting of an electrometer triode followed by a stage of voltage amplification and a cathode follower. The gain of the amplifier is adjusted to unity, so that any voltage change at the input results in an equal change at the output. In addition, this amplifier has a negligible loading effect upon the input of the period amplifier. The current flowing through the operating diode of the period amplifier is the sum of the test current and the ion chamber current as follows:

$$i_t = Ae^{at^2} + I_0$$

making the change in the voltage input to the first stage of the amplifier.

$$v_i = \ln(I_0 + Ae^{at^2})$$

The output of the period amplifier is the derivative of the above, or

$$i_0 = \frac{d}{dt}\ln(I_0 + Ae^{at^2})$$

When the value of Ae^{at^2} becomes large compared with I_0,

$$i_0 = 2at$$

which simulates a linearly rising inverse period.

A reasonable effort has been made to minimize the number of unsafe faults which can occur in the system and to take care of the remaining possible unsafe faults by automatic testing. It has not appeared feasible, however, to apply the automatic tests to the ion chambers and their cables. It has therefore seemed important to make these parts fail-safe. This has been accomplished, in the case of the power chambers, by the following means. The d-c supply for the chamber is a full-wave rectifier with scant filtering. It supplies 300 volts of direct current and 30 volts of 120-cycle ripple to the chamber, which saturates at about 100 volts at full flux. The signal returning from the chamber is therefore composed of a d-c component which is proportional to flux, plus a 120-cycle component which is proportional to interelectrode capacitance within the chamber. As shown in Fig. 3, the 120-cycle signal is separated from the d-c signal after they pass into the safety amplifier, and this 120-cycle signal goes into a 120-cycle amplifier, where its level is raised enough to be able to hold in, after rectification, an auxiliary relay. Any fault in power supply in the ion chamber, whether in the amplifier or in the cables, will cause the relay contacts to open and trip the channel. The one exception to this is an interelectrode short circuit inside the chamber. This type of fault causes a high d-c signal, which simulates a high reactor power and trips the channel, so that it also is a safe fault.

Solid-State Neutron-Flux Measuring System

T. S. GRAY
FELLOW AIEE

W. M. GRIM, JR.
ASSOCIATE MEMBER AIEE

F. S. REPLOGLE, JR.
NONMEMBER AIEE

R. H. SPENCER
ASSOCIATE MEMBER AIEE

FOR CONTROL of a nuclear reactor, direct measurement of the power developed is desirable. The product of the coolant flow rate and the temperature differential of the coolant through the reactor gives a measure of the power developed. But this method of measurement is not fully adequate for control, because the detectors of temperature and flow rate respond too slowly to the rapid changes of power and temperature that can occur in reactors, especially during startup.

To a first approximation, the power developed is directly proportional to the number of fissions per unit time, because each fission is accompanied by the release of a definite amount of energy, approximately 3 times 10^{-14} Btu. Hence the power developed is directly proportional to the rate at which neutrons encounter the fuel—in other words, to the neutron flux within the reactor. Thus, measurement of the neutron flux provides an indirect measure of the power.

Neutrons themselves are not readily measured electrically because they have no electrical charge. But they can be detected indirectly through their ability to cause transmutation of one element into another (for example, boron into lithium) with accompanying release of energetic particles that are capable of causing ionization in a gas or liquid, or temperature rise in a solid.

For control purposes, the ionization phenomenon in a gas has been utilized most generally because the output of a neutron-sensitive gas-filled ionization chamber responds rapidly to changes in neutron flux and the corresponding reactor power.[1,3] To achieve adequate sensitivity, the volume of a neutron-sensitive ionization chamber must be so large as to make the device unduly bulky. Furthermore, even with adequate volume, the electric output current produced at low reactor power levels is in the order of micromicroamperes. Consequently, the insulation resistance of the insulators used within the chamber and in the cables that connect it to the current amplifiers must be extremely high. For these reasons, ionization chambers for reactor control are bulky and fragile, and the associated amplifier equipment is necessarily delicate.

Neutron-sensitive thermopiles that utilize the temperature rise in a solid accompanying transmutation of boron into lithium have also been used for measurement of neutron flux.[3,4] They consist of thermocouple wire junctions in series with alternate junctions coated with boron. Previous designs have the disadvantages that the time response is often several seconds, that the devices are fragile and hence subject to damage by mechanical shock, and that, because air cooling of the hot junctions is used, neutron-flux-to-electric-power conversion is not optimized. This paper is a description of an experimental neutron-flux measuring system that utilizes a rugged neutron-sensitive thermopile designed for rapid response.

The output voltage and internal resistance of this thermopile are so small that amplification by a direct-coupled electronic amplifier is impracticable. Thermocouple outputs are ordinarily amplified through conversion of the d-c signal into alternating current by means of a contactor modulator, or chopper, and subsequent a-c amplification. In the search for a more rugged and longer lived d-c to a-c conversion device without moving parts, early consideration was given to the second-harmonic magnetic modulator. The magnetic modulator operates on the principle that the voltage induced in a coil around a magnetic material contains no even harmonics when the coil is excited by a sinusoidal current source, but that even harmonics do appear in the induced voltage when the excitation is made nonsymmetrical by the superposition of the steady or slowly varying signal current on the excitation.[5] The amplitude of the second harmonic in the induced voltage is essentially proportional to the signal current over a considerable range. The magnetic modulator is well suited for use with detectors having low internal resistance such as the neutron-sensitive thermopile, because the input resistance of the modulator can be adjusted through choice of the wire size and turns in the excitation winding to match the internal resistance of the thermopile and its leads, thereby maximizing the output of the modulator. Furthermore, the second-harmonic modulator can be designed to produce considerable power gain.[6,7] Because of these features, the magnetic modulator is used as the device to accept the thermopile output signal, and simultaneously to modulate and to amplify this signal.

In the particular system evolved, the modulator gain is insufficient to produce the desired output signal level, hence an

---

Paper 57-407, recommended by the AIEE Nucleonic and Radiation Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the 2nd Nuclear Engineering and Science Conference, Philadelphia, Pa., March 11–14, 1957. Manuscript submitted January 25, 1957; made available for printing May 22, 1957.

T. S. GRAY and R. H. SPENCER are with Massachusetts Institute of Technology, Cambridge, Mass.; W. M. GRIM, JR. is with General Electronic Laboratories, Inc., Cambridge, Mass., and F. S. REPLOGLE, Jr. is with Schlumberger Well Surveying Corporation, Ridgefield, Conn.

The authors are indebted to their associates in the Electronic Nuclear Instrumentation Group of the Servomechanisms Laboratory, Massachusetts Institute of Technology, for helpful assistance and suggestions. Dr. A. B. Van Rennes, Prof. T. F. Jones, Jr., C. O. Harbourt, E. J. Kletsky, B. T. Subbotin, R. D. Thornton, and J. K. Watson, in particular, made numerous contributions to the development. The research described in this document was made possible through the support extended the Massachusetts Institute of Technology by the Office of Naval Research and the Navy Bureau of Ships under contract no. *N5ori-07876*.

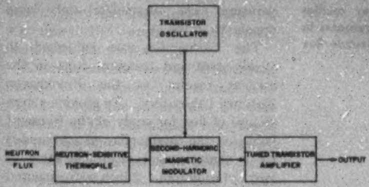

TRANSISTOR OSCILLATOR
NEUTRON FLUX
NEUTRON-SENSITIVE THERMOPILE
SECOND-HARMONIC MAGNETIC MODULATOR
TUNED TRANSISTOR AMPLIFIER
OUTPUT

Fig. 1 (left). Block diagram of neutron-flux measuring system

Fig. 3 (right). Thermopile in clamp

amplifier is included. To provide ruggedness and long life, transistors are used instead of vacuum tubes in the amplifier. It has a narrow passband centered at the output frequency of the modulator, which is the second harmonic of the modulator excitation frequency.

The neutron-flux measuring system is composed, then, of four essential parts: (1) a neutron-sensitive thermopile, (2) a magnetic modulator, (3) an excitation source for the magnetic modulator, and (4) a transistor amplifier. These four components are shown as portions of the block diagram of the system in Fig. 1. A detailed discussion of these parts follows.

Thermopile

The neutron-sensitive thermopile is designed to have a response time much shorter than 1/10 second, so as to be adequate for reactor control. Consideration of the physical dimensions required for response times appreciably less than 1/10 second shows that the dimensions of lumped thermal capacitance and conductance elements make them excessively small and fragile. Consequently, a distributed-constant system is used.

By an elementary analysis, it has been shown that the thermal performance of a simple lumped-constant thermocouple configuration is not basically affected by whether the conduction of heat from the hot junction is first to a cold junction or directly to an ambient sink.[3] It is assumed that the qualitative behavior of a distributed-constant thermopile is likewise independent of whether heat is conducted first to the cold junctions or directly to the ambient sink. It is desirable to design a thermopile so that the heat from the hot junctions is all carried to the cold junctions by way of the electrical conductors, since increasing their thermal conductivity, and correspondingly their electrical conductivity, increases the power that can be delivered by the thermopile to a signal amplifier input which matches the internal resistance of the thermopile.

The considerations of physically rugged construction, small element thicknesses necessary to obtain fast time response, and minimum electrical resistance in a thermocouple circuit lead to the idea of constructing the individual thermocouples in the form of a sequence of three disks, the outer disks being the thermoelectrical materials, and the inner disks being the neutron-flux-to-heat converting material, or heater element, as shown in Fig. 2. These disks are stacked in a long repeating sequence to form a thermopile of mechanically simple and rugged construction.

To achieve fast time response without sacrificing electric signal power output, it is essential that the elements in the stacked-disk thermopile have small electrical and thermal resistance and large area, and that the thermoelectric materials have large thermoelectric power. When the disks are thin, the neutron flux that reaches a particular heater element is reduced by neutron absorption in adjacent heater element disks on either side. Spacer elements *S* in Fig. 1 are added to reduce this geometrical shadowing effect.

Analysis of the transient behavior of the thermopile has been made with the aid of a lumped-constant electrical analog of the distributed-constant thermopile. This analog takes into account the presence of radial heat flow as well as longitudinal flow from the heater elements. The general conclusion regarding the transient output signal from the multielement thermopile is that the output signal will follow qualitatively the form of an initial fast exponential rise and a subsequent slow exponential rise. The time constant of the initial rise is determined by the time constant of a single couple. The time constant of the final rise is a function of the number of couples in the stack and the radial heat conductance. It is reduced as the radial heat conductance is increased, and as the number of thermocouples in the thermopile is reduced.

Preparation of the neutron-flux-to-heat converting material for use as heater elements was a major problem in the thermopile development. Boron was selected as the converting material because it could be fabricated most easily. A material of high boron content and high thermal and electrical conductivity was therefore desired. After unsuccessful experimentation with electroplating processes and hot-metal spraying processes, a method of preparation based on hot pressing a mixture of dry 800-grit boron carbide powder and dry, reduced, precipitated, powdered copper has been evolved. Wafers of this mix are first cold pressed into a thickness of about 0.020 inch and then stacked between oxidized, graphited tantalum disks and placed in a cylindrical graphite press in

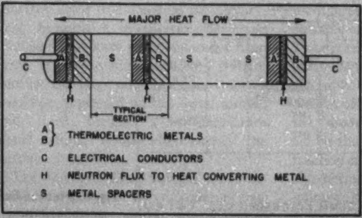


Fig. 2 (left). Stacked-disk-type thermopile

Fig. 4 (right). Thermopile shutter housing parts

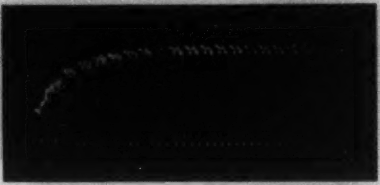


Fig. 5. Cathode-ray oscillogram of thermopile response to a step change of neutron flux

which they are subjected to a temperature of 1,050 degrees centigrade under a pressure of 2,000 pounds per square inch. In this way, 15 wafers each 0.020 inch thick and 3/4 inch in diameter are made in a single hot pressing. These wafers are ground and lapped with a diamond wheel to a thickness of 0.006 inch and then electroplated with an 0.0002-inch-thick layer of copper on opposite faces.

Although bismuth and antimony would provide larger thermoelectric power, bismuth and lead are used as the thermoelectric elements because of difficulties encountered in achieving adequate strength in an antimony layer. Accordingly, bismuth and lead disks 0.016 and 0.066 inch in thickness, respectively, are fused to opposite sides of each boron-carbide and copper disk. This process is carried out under vacuum to provide the best possible bond between the disks so as to achieve good electrical and thermal conductivity. During the vacuum fusion process, copper wafers 0.032 inch thick are fused in turn to the outside faces of the lead and bismuth disks. One hundred such thermopile sections are soft-soldered together to form 2 stacks of 50 couples each. The outer diameter of the resulting cylinders, which are about 8 inches long, is reduced to 5/8 inch by cylindrical grinding between centers.

To avoid development of spurious thermoelectric electromotive forces by action with the electrical conductors attached to the thermopile, it is necessary that the temperatures of the ends of the stack be kept equal. To provide good end-to-end heat conduction and radial heat conduction, the two sections of the thermopile are wrapped with silicone-impregnated fiberglass tape 0.004 inch thick and then clamped in a position parallel to one another by full-length aluminum clamps constructed in a symmetrical manner so as to cancel signals generated in the thermopile itself when a change occurs in external ambient temperature. The thermopile and clamp are shown in Fig. 3.

The thermopile was subjected to steady-state and transient tests in the nuclear reactor at the Brookhaven National Laboratory. To provide a step-change of flux for study of the transient behavior, the thermopile was surrounded by a shutter mechanism consisting of strips of cadmium 1/2 inch wide mounted on an aluminum frame and spaced 1/2 inch apart. In the closed position, the cadmium strips of the shutter had their openings masked by corresponding cadmium strips attached to a stator assembly mounted around the thermopile. The shutter parts are shown in Fig. 4. The shutter could be jerked open by a wire, spring, and trip mechanism in a measured time of 0.007 second.

Measurements indicate that the thermopile, which has a resistance of 3/4 milliohm, produces an output signal of 21 microwatts into a matched resistance load when placed in a position in the reactor for which the unperturbed neutron flux is 3.8 times 10^{12} neutrons per square centimeter per second. Since the thermopile is itself a strongly absorbing body, and its dimensions are comparable with the neutron transport mean free path in the reactor, its presence tends to depress the flux level in its vicinity. Computation of the resulting actual flux level at the thermopile during the test is difficult, but the flux is estimated to have

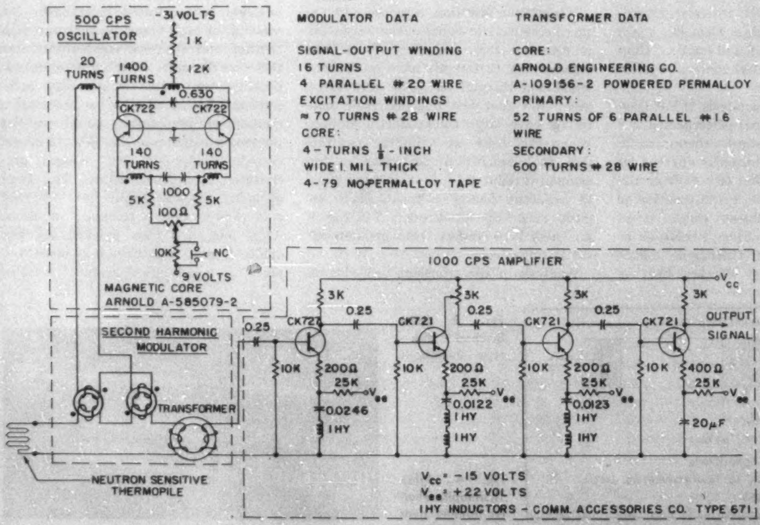

500 CPS OSCILLATOR
-31 VOLTS
1 K
12K
20 TURNS
1400 TURNS
0.630
CK722
CK722
140 TURNS
140 TURNS
1000
5K
5K
100Ω
10K
NC
9 VOLTS
MAGNETIC CORE
ARNOLD A-585079-2
MODULATOR DATA
SIGNAL-OUTPUT WINDING
16 TURNS
4 PARALLEL #20 WIRE
EXCITATION WINDINGS
≈ 70 TURNS #28 WIRE
CORE:
4-TURNS 1/8-INCH
WIDE 1 MIL THICK
4-79 MO-PERMALLOY TAPE
TRANSFORMER DATA
CORE:
ARNOLD ENGINEERING CO.
A-109156-2 POWDERED PERMALLOY
PRIMARY:
52 TURNS OF 6 PARALLEL #16
WIRE
SECONDARY:
600 TURNS #28 WIRE
SECOND HARMONIC MODULATOR
TRANSFORMER
NEUTRON SENSITIVE THERMOPILE
1000 CPS AMPLIFIER
0.25
CK722
CK721
CK721
CK721
3K
3K
3K
3K
0.25
0.25
0.25
10K
10K
10K
10K
200Ω
200Ω
200Ω
400Ω
25K
25K
25K
25K
Vee
Vee
Vee
Vee
0.0246
0.0122
0.0123
20μF
1HY
1HY
1HY
1HY
1HY
Vcc
OUTPUT SIGNAL
Vcc = -15 VOLTS
Vee = +22 VOLTS
1HY INDUCTORS - COMM. ACCESSORIES CO. TYPE 671

Fig. 6. Neutron-flux measuring system

been 25% of the unperturbed value, or 9.5 times 10^{11} neutrons per square centimeter per second.

A sample cathode-ray oscillogram of the output of the thermopile is shown in Fig. 5. From the oscillogram, the time constant of the exponential response of the thermopile to a step function of neutron flux is estimated at 37 milliseconds.

During controlled reductions of the reactor power level in the order of 1 decade of power change in 5 minutes, the thermopile signal showed a tendency to overshoot and even to show negative output signals, amounting to as much as −1.4% of the full power indication. This behavior is attributed to thermal unbalance in the thermopile or in the clamp and shutter assembly. It remains unexplained in full, however, and is a limitation on the dynamic range and minimum detectable flux for which the thermopile is useful.

Magnetic Modulator

An elementary form of the second-harmonic magnetic modulator consists of a toroidal core of magnetic material provided with two windings. One carries a sine-wave current excitation; the other carries a signal current at low frequency or direct current. This second winding is also the output winding. If the signal current is made zero, the device acts very much like a transformer. With an excitation current large enough to drive the core alternately from one saturation level to the other, the output voltage contains the fundamental and many odd harmonics of the input-current frequency. Only very small even-harmonic content is found at the output because of the symmetry of the core hysteresis loop.

If some signal current exists, the core is effectively biased toward one of the saturation levels and even harmonics appear in the output voltage, since the hysteresis loop then appears nonsymmetrical. If a particular even-harmonic voltage amplitude is observed, it is found to be nearly linearly dependent upon the signal-current amplitude. In practice, this modulator would not be satisfactory, first, because a small even-harmonic voltage would have to be separated from large odd-harmonic voltages; second, because the large odd-harmonic voltages would be impressed on the signal source; and, third, because a low-impedance signal source would short-circuit the output.

The first two disadvantages are overcome through use of two modulators connected so that the odd harmonics tend to cancel each other and the even ones tend to enhance each other. To overcome the third difficulty, however, some form of frequency-selective coupling network is still required between the signal source and the combination input and output winding of the modulator.

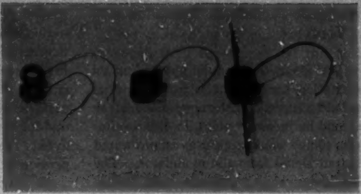


Fig. 7. Assembly of 2-core magnetic modulator

The configuration of the modulator used in the neutron-flux measuring system is shown in Fig. 6. For this type of magnetic modulator, the characteristics which specify the modulator design are the required time constant (or bandwidth) and input resistance. Once the time response is specified, the relationship between gain and excitation frequency is fixed. Further, the time constant also specifies the core cross-section area. The modulator for this system was designed with the expectation that an input resistance of 10 milliohms and a time constant of 25 milliseconds would be required. Actually, the resistance of the thermopile is smaller than this value, but the leads to the thermopile were expected to make the total output resistance of the thermopile approach this value. Mo-permalloy cores are used and the time-constant specification requires that a cross-sectional area of 2.6×10^{-3} square centimeters be used. The tape-wound core used is constructed of 4 turns of 1/8-inch-wide 1-mil-thick tape with an average diameter of 0.5 inch. Calculation indicates that 16 turns of no. 14 wire are needed for the input signal winding. As no. 14 wire is too heavy for convenient winding, 4 parallel strands of no. 20 wire are used instead.

A transformer with its primary winding in series with the thermopile and the modulator signal winding serves as a suitable coupling network in this system. The primary impedance of the transformer is low for the input signal and considerably higher for the second-harmonic output of the modulator. Thus, most of the modulated output signal appears across the transformer while only a small fraction of the input signal is lost across it. Furthermore, the transformer provides voltage step-up for the input to the subsequent amplifier.

In addition to having proper magnetizing inductance and winding resistance, the transformer must be constructed of core material having negligible hysteresis. If hysteresis is present, the nonsymmetrical loop resulting after a large signal is applied and then removed, will cause the transformer itself to act as a modulator, excited by the residual fundamental output of the 2-core modulator. It hence tends to produce a spurious even-harmonic output, dependent upon the previous history of the modulator system. This requirement dictates the use of a magnetic core with an air gap or a powdered ferromagnetic core. In the system, an Arnold Engineering Company no. *A-109156-2* powdered permalloy core is used.

The details of the actual numbers of turns used on the specific cores are shown in Fig. 6. The construction is illustrated in Fig. 7. Here, the actual modulator is shown in three stages of construction. On the left, the two cores are shown with the excitation windings applied; in the middle, the two cores are placed together with a rubber separator; and on the right, the signal winding is shown wrapped around the two cores.

The magnetic field intensity corresponding to the minimum detectable signal of the modulator is 10^{-5} oersted. For this reason, the entire modulator, together with its transformer, is placed in a magnetic shield consisting of triple-nested shield cans. This shield is effective in reducing the effects of stray magnetic fields on the modulator. The characteristics of the magnetic modulator are listed in Table I.

Table I. Magnetic Modulator, Including Output Transformer

Quantities	Value
Input resistance	0.01 ohm
Minimum input signal	0.03 microvolt
Maximum input signal	30 microvolts
Dynamic range	1,000 to 1
Time constant	0.025 second
Output resistance	300 ohms
Power gain (to matched load)	20
Voltage gain (to matched load)	1,500
Minimum output signal	45 microvolts
Maximum output signal	45 millivolts

Magnetic-Modulator Excitation Source

The magnetic modulator requires an excitation source that is nearly free of second-harmonic distortion, since stray second-harmonic content in this excitation source would appear as an unwanted output signal from the modulator. The modulator requires approximately 3 milliwatts of power at the operating frequency of 500 cps (cycles per second), with a maximum of 0.01% second-harmonic output.

The oscillator that provides the excitation for the modulator in the system shown in Fig. 6 consists of two transistors which alternately switch current from one side to the other of a tuned center-tapped transformer having a powdered permalloy core. The transformer has one winding to provide the transistor switching voltages and another to provide the output. Low second-harmonic content results when each transistor conducts for exactly 1/2 cycle. The circuit design is such that the conduction periods tend to be equalized even though the transistors are slightly dissimilar.[9]

Transistor Amplifier

The purpose of the transistor amplifier is to raise the signal level from the magnetic modulator to a useful level. Consequently, the output range and characteristics of the magnetic modulator must be taken into account in the design of the amplifier. At the lower extreme, the 45-microvolt 1,000-cps signal must be amplified in the presence of about 1 millivolt of extraneous odd-harmonic voltages. The presence of these residual signals requires the use of a tuned amplifier. Furthermore, the input stage of this tuned amplifier must not generate 1,000-cps signal through second-harmonic distortion of the residual 500-cps signal appearing at the input.

The basic tuned stage of the amplifier consists of common-emitter degenerative stages. In each stage the degeneration is made frequency-selective by the use of an *LC* series-tuned circuit between the emitter and ground, tuned to a frequency near the second-harmonic frequency. At the resonant frequency, the impedance of this circuit is small, making the gain of the stage relatively large. At frequencies different from the resonant value, the impedance becomes large, reducing the gain of the stage. The action is very much like that in a cathode-degenerative vacuum-tube amplifier having a series-tuned circuit between the cathode and ground. In addition, the cathode degeneration maintains the second-harmonic distortion small so as to avoid the spurious signal mentioned previously.

To provide a sufficiently fast fall-off of gain for frequencies outside the required passband, and to keep the gain within the required passband sufficiently level, three "flat-staggered" tuned stages are used. The fourth stage of the amplifier is an untuned stage. Gain adjustment of the amplifier is provided as indicated in the circuit diagram of Fig. 6. The dynamic range is limited at the lower end by noise and at the upper end by nonlinearity in the transistors.

The tuning inductances are the same type of powdered-permalloy cores used in the modulator-excitation oscillator. Although these cores are temperature stabilized as are the tuning capacitors, some slight change in tuned frequency can occur with temperature. Frequency drift caused by temperature, however, should be the same in both the excitation oscillator and the amplifier. This feature keeps the system operative over a wider temperature range than would otherwise obtain. The over-all characteristics of the transistor amplifier are listed in Table II.

System Characteristics

Experimental data on the behavior of the over-all system are not available. On the basis of the performance of the individual components given above, however, the over-all system may be expected to have the following characteristics:

Maximum sensitivity....................

$$\frac{9\times 10^{-11}\ \text{rms amplifier output volts}}{\text{neutron/square centimeter/second}}$$

Response time..............45 milliseconds

Rated maximum neutron flux............

4×10^{13} neutrons/square centimeter/second (unperturbed)

Implicit in these results are the assumption that (1) the total internal and lead resistance of the thermopile matches the input resistance of the modulator, (2) the over-all response time is the square root of the sum of the squares of the thermopile, modulator, and amplifier response times, and (3) when rated maximum neutron flux is measured, attenuation is introduced to avoid overload of the modulator and amplifier. Omission of the cadmium shutter and better matching of the thermopile, modulator, and amplifier might be expected to produce considerable improvement in the sensitivity of the system.

Table II. Transistor Amplifier

Quantity	Value
Voltage gain	0 to 250
Passband	±40 cps
Tuned frequency	1,000 cps
Input resistance	4,000 ohms
Output resistance	3,000 ohms
Minimum detectable input signal	5 microvolts
Dynamic range	3,000 to 1
Maximum output voltage	3 volts
Gain at 500 cps and 1,500 cps	Down by a factor of >1,000 over that at 1,000 cps

The output of the amplifier is adequate to excite existing electronic drivers for control-rod actuators. Consequently, this system is believed to be a step toward more reliable and longer-lived reactor control equipment.

References

1. INTRODUCTION TO NUCLEAR ENGINEERING (book), Richard Stephenson. McGraw-Hill Book Company, Inc., New York, N. Y., 1954, Fig. 7–9, p. 286, Fig. 7–10, p. 288.

2. A NEUTRON-SENSITIVE IONIZATION CHAMBER WITH ELECTRICALLY ADJUSTED GAMMA COMPENSATION, H. S. McCreary, Jr., R. T. Bayard. *Review of Scientific Instruments*, New York, N. Y., vol. 25, no. 2, 1954, pp. 161–64.

3. THE DESIGN AND CONSTRUCTION OF BORON COATED THERMOPILES FOR USE IN NEUTRON FIELDS, G. Barbaros, J. Farr, J. Kuranz, R. J. Moon, L. M. Rieser, G. W. Shaeffer, T. Wartek. *Document No. AECD-2975; CP-2697*, Technical Information Service, Atomic Energy Commission, Oak Ridge, Tenn., Feb. 7, 1945 (declassified June 2, 1950).

4. NEUTRON FLUX MEASUREMENTS WITH THERMOPILES, A. C. Lapsley. *Nucleonics*, New York, N. Y., vol. 11, no. 5, May 1953, pp. 62–64.

5. MAGNETIC MODULATORS, E. P. Felch, V. E. Legg, F. G. Merrill. *Electronics*, New York, N. Y., vol. 25, no. 2, Feb. 1952, pp. 113–17.

6. THE DESIGN OF A PRACTICAL D.C. AMPLIFIER BASED ON THE SECOND HARMONIC TYPE OF MAGNETIC MODULATOR, S. W. Noble, P. J. Baxandall. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. 2, no. 70, Aug. 1952, pp. 327–44.

7. DESIGN CRITERIA FOR LOW-LEVEL SECOND-HARMONIC MAGNETIC MODULATORS, E. J. Kletsky. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. XI, Oct. 3–5, 1955, pp. 890–906.

8. A STACKED-DISK NEUTRON THERMOPILE, F. S. Replogle, Jr. *Technical Report No. 5*, Electronic Nuclear Instrumentation Group, Servomechanisms Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., Mar. 15, 1955; *ASTIA No. AD-61-542*, Armed Services Technical Information Agency, Dayton, Ohio.

9. A SYMMETRICAL TRANSISTOR OSCILLATOR WITH LOW SECOND-HARMONIC DISTORTION, W. M. Grim, Jr. *AIEE Transactions*, vol. 74, pt. I, 1955 (Jan. 1956 section), pp. 767–70.

Some Operational Problems of the Nuclear Power Plant

A. R. JONES
ASSOCIATE MEMBER AIEE

R. L. WITZKE
FELLOW AIEE

IN PLACE of the boiler, as found in the steam power plant today, a nuclear-fueled steam generator will be applied in many instances in the years immediately ahead. In fact, the atomic plants already in design exceed 1,000,000 kw capacity. The industry has been so deeply involved in the evaluation of the economic feasibility of nuclear power that very little attention has been devoted to the problems which are to be expected in the operation of these plants. In general, this phase of the problem has been dismissed by one of two statements: (a) The nuclear plant should be only slightly more difficult to operate than the conventional plant. (b) The nuclear plant will require about the same kind of operating crew as will the conventional plant. The authors subscribe to both of these general statements, but they also believe the time has come to look more closely at the details of nuclear plant operation. Glasstone[1] indicates less than one mill per kilowatt-hour variation in the estimated ultimate total cost of power between five reactor types. These are recognized to be rough estimates, but they do indicate that rather small factors of cost, such as operation and maintenance, may strongly influence the final selection of reactor type.

It is not feasible, unfortunately, in a single general paper such as this, to examine the operational problems of all these reactor types. In fact, the authors can do little more than point out a few of the problems to be dealt with in one reactor type. If this paper can provide the stimulus for further reports which set forth the problems of each type, it will have served an important purpose. The authors would implore the industry to prepare, as soon as possible, a series of reports on the studies to define the operational costs of the various reactor types. The papers by Love and Oldham[2] represent a beginning on this program.

The pressurized water reactor has been selected for discussion in this paper for several reasons. Aside from the authors' greater familiarity with this type, much more information is generally available on the plant. An even more important reason is to be found in the fact that of the seven large nuclear plants planned for construction in the next 6 years, three will utilize pressurized water reactors, and a fourth is a modification thereof.

The normal performance of a nuclear plant is generally accepted as excellent for power production. Most of the problems discussed herein have to do with auxiliary systems and emergency conditions. Before these are discussed in detail, it will be helpful to develop a common concept of the plant under consideration.

The problems are grouped under the following operational headings: start-up, normal operation, shutdown, maintenance, refueling, and manpower.

The Reactor Type

The pressurized water reactor (see Fig. 1) derives heat energy from the fission of uranium and other heavy atoms. These are maintained in a fuel structure of solid rods or plates, inside a reactor or pressure vessel. The energy is transported in hot water, at approximately 2,000 pounds per square inch absolute and 500 to 600 F, to the steam generator where water, in a separate system, is boiled to produce steam for the turbine. Depending upon the plant size, one or more heat transfer loops and steam generators may be utilized to remove the heat from a single reactor. The water in these loops is maintained with less than 2 ppm (parts per million) of impurities by a purification system. As mentioned previously, a system is provided to pressurize the water to 2,000 psia. A makeup and drain system, with connections to a water supply of high purity, and with a waste-disposal system, is required to keep the appropriate water level in the system. When heat energy is not required to drive the turbine, heat must be removed from the reactor by the shutdown cooling system. Reactor control and radiation monitoring of a high reliability are required for the safety of the plant.

Start-up

The start-up problems depend somewhat on the previous history of plant operation. For example there are certain special precautions to be observed in the initial startup of a new plant. This paper makes no attempt to deal with these. Similar problems arise when a new fuel structure, of modified design for taking advantage of developmental progress, is put into operation. Such operating periods are assumed to be carefully supervised by the equipment manufacturer and are of such infrequent occurrence as to warrant no more than mention here. The first case is encountered only once in the plant lifetime and the second can be expected at perhaps 1 to 5-year intervals in the early plants, and less frequently with time.

The startup of a cold plant can be expected at irregular intervals and should be accomplished by the normal operating crew. It is therefore of more immediate and general interest. For the moment assume that the primary and secondary systems, of approximately 300 to 600 gallons per mw (megawatt) capacity, are filled with water of the desired purity (approximately 1 to 2 ppm).

The first requirement of start up is of course the check out of instrumentation and components to insure proper function. The items to be checked, and the procedure, will be peculiar in many cases for the nuclear plant, but the problem is not essentially different from that encountered in any other plant.

Cranking power to drive the plant auxiliaries must be supplied. For the electrical utility application this usually means energizing the auxiliary power bus from the system. For isolated units a diesel generator may be used to provide this power. Available data[3] indicate that the auxiliary requirements are about 8 to 9% of rated output. A sizeable fraction of this, at least one third and perhaps two thirds, is needed during startup.

Oxygen scavenging, described by Love,[2] may require appreciable time (up to 4 hours) before actual startup, if the system has been refilled with supply water. The scavenging is accomplished by the addi-

---

Paper 57-396, recommended by the AIEE Nucleonic Radiation Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the 2nd Nuclear Engineering and Science Conference, Philadelphia, Pa., March 11-14, 1957. Manuscript submitted December 17, 1956; made available for printing May 22, 1957.

A. R. Jones and R. L. Witzke are with Westinghouse Electric Corporation, Pittsburgh, Pa.

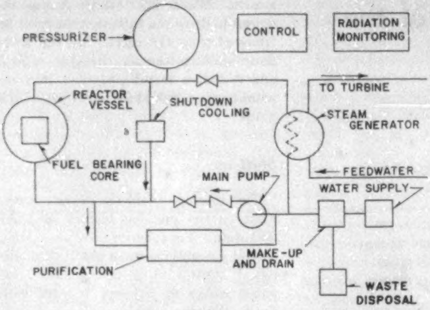


Fig. 1. Schematic of pressurized water reactor plant

tion of hydrazine. For this the system is pressurized just enough to provide the required net positive suction pressure of 100 to 200 pounds per square inch at the main coolant pump intakes, either with the pressurizer or with the charging pumps. These pumps are then operated to circulate the hydrazine solution and permit the hydrazine oxygen reaction to take place. During the oxygen scavenging, and indeed throughout the start-up procedure, all main coolant valves should be open.

Design practice is such that the reactor will be shut down before full insertion of the control rods is reached. However, the operating practice is to take advantage of this extra margin of available rod insertion, as a safety factor on shutdown. Therefore the control rods must be pulled for a considerable distance before the reactor reaches criticality and begins to produce power. In order to avoid accidental overshoot, this rod withdrawal is done at a very slow rate, and a period of 1/2 hour may be so consumed. This approach to criticality should be made only when the main coolant pumps are operating. Likewise, the electric heaters in the pressurizer should be operating, so that a steam bubble space is available in the system for water expansion. Once criticality is obtained, the plant is ready for warm up.

This operation requires careful coördination with the pressurizer, and for two reasons. First, the temperature of the steam bubble in the pressurizer must exceed the maximum fuel surface temperature to prevent boiling and burnout. Second, the required net positive pressure must be maintained at the pump suction to prevent cavitation. Maximum pressurizer heater power demand occurs at this time. The actual rate of system warm up is limited in order to avoid excessive thermal stresses in the reactor vessel walls. One hundred to 200 degrees per hour are permitted. The allowable rate of rise is influenced by the material selection, and is inversely proportional to the wall thickness. The rise from room temperature to 500 F will require, then, about $2^{1}/_{2}$ to 4 hours.

During the warm-up period the density of the water will drop from 62.4 to 49.6 pounds per cubic foot[3] at 500 F. To accommodate this expansion it is necessary to drain from 30 to 60 gallons per mw capacity from the system. The purity of this water is such that it is desirable to store rather than waste it.

Up to 8 hours may be required for a cold plant start up. This compares with 5 to 6 hours for a conventional plant. The time required for a hot plant restart is short, perhaps one to two hours, and compares favorably with that for a conventional unit.

Normal Operation

The operation of an atomic plant in the normal power range appears an exceedingly routine task. Design practice is to produce a reactor with a negative temperature coefficient, making the unit self-regulating to a large degree. The control system is automatic in most cases, and is normally designed to assist the inherent self-regulating characteristic of the turbine-generator unit without adjustment by the operators.

The reactor plant can be designed to follow the load swings demanded by system requirements. This means that the normal turbine limitations are the main restrictions on plant operation. The plant is capable of sustained operation at any fraction of full load above the minimum (5 to 10%) set by turbine-generator restrictions.

The radioactivity, present throughout the primary system during operation, prevents approach for maintenance. Of course, this does not apply to many of the auxiliary systems. The design, then, must be such as to minimize the need for attention for long periods if the plant is to meet the needs of the electrical utility application. Minimization of maintenance is obtained by the use of such a component as the canned motor pump, in which bearings are lubricated by the pumped fluid and shaft seals, and where packing boxes are eliminated by hermetically sealing the entire rotating portion of pump and motor inside the system.

These comments do not mean that the plant should be left unattended. Most of the auxiliary systems can, and possibly should, be manually controlled, to take advantage of the operating personnel who must be maintained on the payroll. This will minimize the capital expenditure for automatic control and self-operating systems. Those auxiliary systems not required to operate continuously should be designed to be as conventional as possible for the same reason.

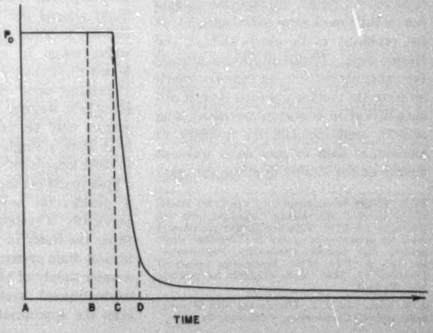


Fig. 2. Typical curve of heat after shutdown

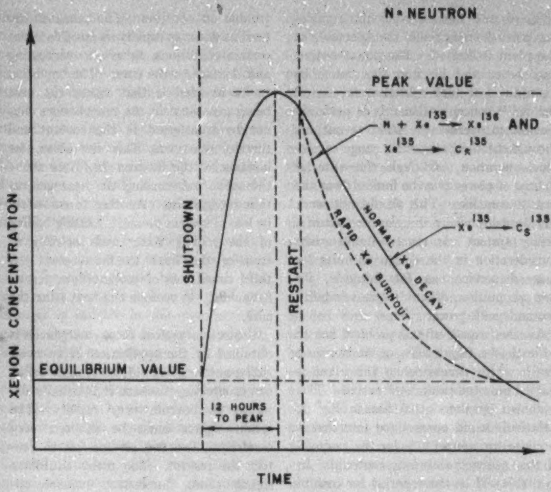


Fig. 3. Typical curve of transient xenon after shutdown.

Shutdown—Normal

The production of heat energy can be stopped in three distinct ways. In certain cases the system power demand will be zero for short periods. In such instances it is desirable and usually feasible to keep the reactor at temperature and critial without producing appreciable power. The auxiliaries are kept in operation during such down time, and the power required may be taken either from the main turbine or the system. In other cases the power demand will be zero for extended periods of time, and then it will be desirable to make the reactor subcritical by complete insertion of the control rods. The plant can then be cooled down and will require a minimum attention during the shutdown period. The same thermal stress considerations limit reactor cooldown as those concerning startup. In this latter case the restart corresponds to that described previously. These two are routine operations, and are discussed in this section. The third is an emergency procedure, and will be covered in the following section.

Regardless of the way in which the plant is shut down, two inportant problems must be solved. One of these is the removal of heat which the fission products continue to release even after the chain reaction is arrested. This condition is indicated in Fig. 2. A reactor, operating at power *Po*, is shut down by inserting the control rods at time *C*. Neglecting, for the moment, the transient condition immediately following *C*, the heat release after shutdown[4] is found by the use of the equation:

$$\frac{P(t)}{Po} \approx 0.0315t^{-0.2}$$

where

10 seconds $<t<$100 days
$P(t)$ = heat power release as a function of time after shutdown
Po = heat power release prior to shutdown
t = time after shutdown in seconds
at $t=10$ seconds

$$t=10^{-0.2}=\frac{1}{1.585}=0.63$$

$$\frac{P_{10}}{Po}=0.0315\times0.63=0.0198$$

at $t=100$ seconds

$$\frac{P_{100}}{Po}=0.0125$$

Although this is not a large amount of power, it would be sufficient to damage the reactor core if not removed. The plant design must be worked out to assure proper function after shutdown. The plant layout is made to insure circulation of water in the primary loop by thermal siphon action. Even in the extreme case of loss of auxiliary power to the main coolant pumps, the heat after shutdown is thereby removed.

When the loops are to be taken out of service for maintenance or refueling, the heat must be removed with a separate auxiliary shutdown cooling system (see Fig. 1). This system consists of a pump and heat exchanger, along with suitable valves and piping. It may be manually controlled, and is not required during plant operation at power.

The second important problem after shutdown is the build-up of poisons.

When the uranium atom fissions, the two parts form atoms of different chemical materials. As the result of this process, many new materials are formed in the core. Some of these present large target areas for the absorption of neutrons. The worst offender in this respect is xenon (Xe). A relatively small amount of Xe is formed directly, but over 5% of the fissions produce tellurium (Te). By the process of radioactive decay, this element decays to iodine (I) and then to Xe:

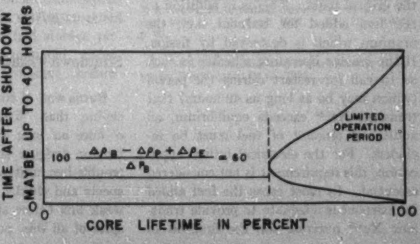
$$Te^{135}\rightarrow I^{135}\rightarrow Xe^{135}$$

During each of these changes, a high-energy electron, a beta particle, is emitted As a result of these processes, an appreciable amount of Xe is present in a reacting core.

The magnitude of the neutron absorbing characteristic of Xe^{135} is indicated by its target area of 3,200,000 barns (1 barn $=10^{-24}$ square centimeters.) By comparison, cadmium (Cd) is considered

Fig. 4. Illustration of xenon limitation on reactor operation

$\Delta\rho_B$ = Reactivity inserted for burnout, 0.05
$\Delta\rho_P$ = Reactivity required to overcome peak poisons, 0.07
$\Delta\rho_E$ = Reactivity required to overcome equilibrium poisons, 0.05



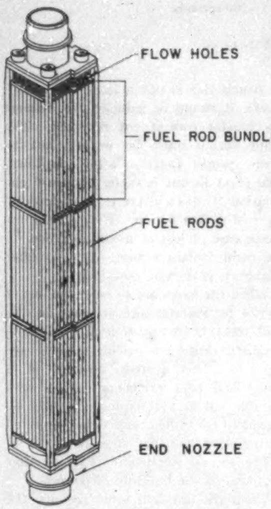


Fig. 5. Reactor fuel assembly

a good control rod material with a target area of 3,300 barns.

Xe^{135} disappears from the reactor in two ways. It decays naturally to cesium (Cs) or, absorbing a neutron, becomes Xe^{136}, which is an isotope of only 0.15 barns target area. When the chain reaction is in progress, both of these processes are operative to limit the build-up of Xe^{135} to a definite equilibrium value. Immediately following shutdown from power, the Xe^{135} concentration rises sharply as the Te^{135} and I^{135} decay into Xe^{135}, and only the Xe^{135} decay into Cs^{135} operates to remove it. Gradually these decay actions are carried toward completion, and the Xe^{135} concentration again falls off, as shown by the solid curve in Fig. 3.

The equilibrium Xe^{135} must be compensated by the addition of fuel beyond the critical mass. This is in addition to the fuel added for burnout, i.e., the uranium which is destroyed by fission. If the reactor operating schedule is such as to call for restart during the period (which may be as long as 40 hours) that transient Xe^{135} exceeds equilibrium, an additional amount of fuel must be included. For the electrical utility application, this requirement is not considered essential. In most cases the fuel added for burnout is adequate to provide transient Xe^{135} override for the first 50 to 75% of core life. During the remaining portion of core life, the operation of the plant is limited. The power output and hence the neutron flux must be maintained at such a level as to control the Xe^{135} concentration. It is perfectly feasible to operate the plant at outputs throughout the permissible range of turbine operation, although the rate of change of power may be limited near the end of core life. This should not work any hardship where the plant is part of a utility system. It must be taken into consideration in the design of units for isolated service, as for example, for ship propulsion, where maneuverability demands wide power swings.

Another aspect of this problem has to do with the rapid burnout of transient Xe^{135}, which occurs when the plant is started near the peak Xe^{135} period. This is shown by the dotted line in Fig. 3. Relatively rapid control rod insertion is required, to compensate for the removal of this neutron absorbing material. In fact, this will be the criterion for control rod speed in most cases.

For the purpose of illustrating the xenon control problem, simplified calculations were made on a thermal reactor assumed to be designed to generate 150 million Btu per hour of heat energy with an average neutron flux of 1.6×10^{13} neutrons per square centimeter per second. The limitation on operation is shown in Fig. 4. During the first 60% of core lifetime, the poisons impose no limit on operation. During the final 40% of core lifetime, a period of 10 hours or less, decreasing to zero at the end of life, is available for normal restart. This curve, the limited operation portion of core life, is influenced almost directly by neutron flux, and hence by power level. This curve has the same shape as the peak xenon curve in Fig. 3, except that the base is the equilibrium value.

During the limited operation period the plant can be restarted at reduced temperature. If a temperature coefficient of reactivity of $-5\times10^{-4}\Delta\rho$ F is assumed and if $\Delta\rho_P-\Delta\rho_E-0.02$, the temperature reduction required at end of life is $0.020/(5\times10^{4})=40$ F.

Shutdown—Emergency

Burns was, of course, perfectly right in saying that "The best laid schemes o' mice an' men gang aft a-gley." The reactor designer struggles to produce a trouble-free unit, and the operator inspects and checks for every conceivable weak link before starting the plant. In spite of all this, power plants will cause trouble on occasion. The plant should be shut down as rapidly as possible under certain conditions, to avoid overheating and damage to the core. The condition to be avoided is that where the heat being generated in the reactor core cannot be transferred to the coolant and thereby removed. This can occur, for instance, if the neutron flux rises above the safe level, causing the heat generation rate to rise. Another cause would be loss of system pressure, causing boiling of the cooling water and inability to transfer the heat to the coolant. A third cause, loss of coolant flow, results in inability to remove the heat from the core.

Protection against these emergencies is obtained by the application of an overriding automatic shutdown system, called scram circuitry because it is usually designed to operate very rapidly. The scram signals must be selected very carefully. Too few scrams fail to protect the reactor. Too many shutdown signals cause unnecessary expense, operating effort, and shutdowns. In general, three parameters, neutron flux, low pressure, and low flow, are adequate to protect the plant.

The period of interest in these shutdown situations is the first few seconds following the occurrence of the difficulty, which can be taken as time *B* in Fig. 2. Some time must elapse before the reactor is shut down at time *D*. The case of loss of power to the main coolant pumps is a good illustration. The inertia of these coolant systems must be examined in each design, and in some cases additional inertia must be added to insure that the transient heat energy is safely removed. Jones and Baldwin[3] have discussed various means for providing this inertia.

The design of the reactor system should be predicated on fail-safe shutdown on loss of auxiliary power. However, it is very desirable that the radiation detection equipment, for example, as well as certain other auxiliary loads, be continuously supplied with power. These loads are normally connected to a "vital bus," which is supplied from a combination of sources that includes the auxiliary bus, a battery, and even a diesel generator.

Maintenance

The maintenance of a nuclear power plant differs from that associated with a conventional plant because the problem of radiation is introduced. When the plant is in operation the primary coolant loop cannot be approached for maintenance. Special considerations must be observed in the design, to eliminate the necessity for frequent attention.[5] Corrosion products, which would become highly radioactive, are removed to minimize the activity in the loop. Valve and pump seals and packings are eliminated by the use of hermetically sealed units. Provisions are made to valve off malfunctioning components pending scheduled shutdown.

In plants above 20,000 kw rating, more than one heat transfer loop and steam generator would be utilized. In such units, a major leak in the heat exchanger is isolable in this manner. Of course, the unit must be operated at part load pending maintenance.

All maintenance of a pressurized water nuclear power plant can be performed in complete safety by the strict observance of fairly simple ventilating, waste disposal, and health physics practices. Careful training and supervision of maintenance personnel is of great importance. For example, crew members should be trained to use radiation monitoring equipment and protective clothing, hats, gloves, coveralls, boots, etc. Covering a floor with butcher and blotter paper to catch radioactive drippings can save endless hours of decontamination work. Control of tools and equipment prevents the spread of contamination.

Appreciable quantities of water are required for rinse prior to any maintenance of systems which handle radioactive coolant. In those instances where radioactive particles have plated out on the surface of the malfunctioning part, an acid rinse followed by water rinse can materially reduce difficulty, time, and manpower. Such a technique would be particularly applicable if fuel element failure had been experienced.

The normal maintenance schedule should coincide with the fuel reloading schedule, which is discussed in the following section.

Refueling

The refueling of a pressurized water reactor usually must be done by taking the plant off the line, cooling down, and opening the reactor vessel. The present procedure allows fuel element design for periods of continuous operation of approximately one year. Intensive testing and developmental work are being carried on to extend this period. While there is very little operating experience to draw upon, present estimates indicate that a well-trained crew can reload a plant with a total down time of one week. As mentioned in the previous section, the maintenance of the primary coolant system should be planned for and performed during this period.

By designing the reactor compartment so that the area above the reactor can be flooded with water, the refueling operation can be carried out without the use of heavy shielding coffins. The very adequate visibility obtained with water makes such an operation desirable. The cost and technical complexity of the manipulating equipment are minimized, as is the time required for the operation.

A large portion of the reloading period is taken up in the removal of shielding sections, auxiliary power and cooling connections, and the reactor vessel head. The present practice is to weld the closure with either strength or seal welds. However, experience indicates that gasket closure is adequate and should be adopted to reduce cost and handling time. Several designs of quick opening heads are under study, and some further advancement can be expected in this area.

The handling of the heavy reactor vessel head requires that the jigs and fixtures be carefully designed to insure against damage of control mechanisms and core during removal and replacement.

After the spent fuel charge has been replaced, and the plant is in operation, the crew can devote its attention to the preparation of the spent fuel for shipment to the reprocessing center. Underwater tools can be utilized to remove non-fuel bearing parts such as the end nozzles (see Fig. 5). After cooling, the fuel-bearing assemblies are loaded for shipment into lead coffins which are approximately 10 inches in wall thickness, and sometimes include a cooling system. Since the water in the fuel storage pit will probably be contaminated, these coffins will require scrubbing before shipment, to remove radioactive material from their outer surfaces.

The waste disposal will be most heavily loaded in the refueling and maintenance period, and storage for subsequent disposal should be considered in the design.

Manpower

Conclusions as to the manpower requirements for an atomic plant are hard to reach. Lyman[6] indicates that 129 positions are planned for the plant at Shippingport, Pa. He further indicates that at least 26 of these will be involved in the developmental portion of the project, and estimates that the final operating crew can probably be reduced to 81. He compares this with a complement of 66 men for a conventional plant of the same rating. These data point up the extra effort needed to get a new plant type in operation, and the extra personnel needed to take and analyze the operational data obtained from the first plants. The value of such data to the industry is obvious, and one need not dwell on the desirability of obtaining and reporting it.

During the early years of the nuclear power program the operating crew will require special training and skills. All special requirements will become commonplace as the indsutry develops, and should be regarded as developmental in nature. It seems to the authors that it would be unfair and unwise to handle the charges for such efforts in any other way.

References

1. Principles of Nuclear Reactor Engineering, a book, **S. Glasstone.** D. Van Nostrand Company, Inc., Princeton, N. J., 1955, p. 780.

2. Getting PWR on the Line, **L. R. Love, G. N. Oldham,** *Nucleonics*, New York, N. Y., July 1956, pp. 27–29.

3. New Problems in Auxiliary System Design—Supercritical-Pressure and Nuclear Plants, **A. R. Jones, C. J. Baldwin, Jr.,** *AIEE Transactions*, vol. 76, pt. III, June 1957.

4. Fission Product Decay Gamma Energy Spectrum, APEX-134, **John Moteff.** Aircraft Nuclear Project, General Electric Company, Cincinnati, Ohio, June 1953.

5. Maintaining Nuclear Reactor Plants, **R. E. Crews, M. L. Yadon.** *Power*, New York, N. Y., Aug. 1956, p. 104; Sept. 1956, p. 97; Nov. 1956, p. 90.

6. The Industrial Atom. *Report no. TID-8010*, Office of Technical Services, Department of Commerce, Washington, D. C., 1955.

Electric Nonlinear Computation Performed by Linear Elements

M. S. M. ABOU-HUSSEIN
ASSOCIATE MEMBER AIEE

THIS paper explains how nonlinear computation including inversion, determining powers, positive and negative roots, logarithms and antilogarithms; and obtaining the hyperbolic functions may be made by using simplified linear circuits. Such circuits may be fed with d-c or a-c power for the different computations mentioned, except for direct multiplication and obtaining the hyperbolic functions where the circuits must be fed with a-c power. In all cases the operation of the circuits is simple, quick, and systematic, and ordinary measuring instruments (ammeters and voltmeters) are used.

The importance of developing such circuits appears when an analog computer, using linear elements, is needed to solve complicated equations involving the different mathematical operations mentioned. Under these conditions, the given circuits will necessarily form basic repeated units of the analog computer.

Theory

Linear computations including direct division and multiplication are explained first, followed by linear circuit representation of the mentioned nonlinear computations.

Division (A/B)

It appears from the basic circuit shown in Fig. 1 that when 1 ampere is adjusted to flow through the lower circuit the result is

$$B=R+1 \qquad (1)$$

The current flowing through the upper circuit will be equal to the voltage drop across the 1-ohm resistance of the same circuit and is equal to

$$A/(R+1)=A/B \qquad (2)$$

i.e., equal to the required division.

It is important to notice that the division is performed without measuring A or B. The resistances R are two identical, isolated, and mechanically coupled resistances. This circuit forms directly or indirectly the basic repeated unit in the following representations.

Multiplication (AB)

Considering Fig. 2, adjusting the resistance of the lower circuit to the shown value of $B(R+1)$ to allow 1 ampere to flow through the same circuit, then

$$B(R+1)=1 \qquad (3)$$

and the voltage appearing across the 1-ohm resistance of the upper circuit thus is

$$A/(R+1)=AB \qquad (4)$$

A means of performing this multiplication experimentally is shown in Fig. 3.

Note that equation 3 necessitates that one of the multiplied quantities B should be less than unity; this limits the circuit operation to such cases only. This circuit may be used, however, to multiply any two scalar quantities A and B if the ammeter reading is adjusted to n amperes instead of 1 ampere. In this case equation 3 becomes

$$B(R+1)=1/n \qquad (5)$$

and the voltage appearing across the 1-ohm resistance of the upper circuit of Fig. 2 is nAB.

Nth Order Inversion

The nth-order inversion B^{-n} is considered where n is a positive integer. Fig. 4 shows how the first-order B^{-1} and second-order B^{-2} inversions of B may be obtained by linear elements. There are three circuits with the indicated values of the applied voltages and connections.

If the ammeter of the lower circuit is adjusted, using R_1, to read 1 ampere, then the voltage appearing across the 1-ohm shunt resistance of the medium circuit will be $1/B$, the first-order inversion.

Similarly, if the ammeter of the medium circuit is adjusted to read 1 ampere, using R_2, then the voltage across the 1-ohm shunt resistance of the upper circuit will be B^{-2}, which is the second-order inversion.

Accordingly, if the total number of circuits used is $(n+1)$, then B^{-n} is finally obtained across the 1-ohm shunt resistance of the $(n+1)$ circuit when the n ammeters used in all circuits except the $(n+1)$ circuit read unity.

It remains to stress that the accuracy of such circuits will be good enough for computation purposes if, in all circuits, the value of the resistance $R_{(n-1)}$ is neglected with respect to the value of the resistance R_n.

Nth Order Power

The nth-order power B^{+n} is considered where n is a positive integer. By tracing the different circuits given in Fig. 5, the third-order power B^3 of B is obtained using the basic circuit of Fig. 1 as a repeated unit. The procedure for obtaining the nth-order power of B is as follows:

1. Apply the voltages B and 1 volt where indicated. Adjust R_1 to let 1 ampere flow through A_1. The voltage appearing across the 1-ohm shunt resistance of the 1-volt circuit is B^{-1}.
2. Divide B by B^{-1} using the R_2 circuits. When A_2 reads 1 ampere, the voltage appearing across the 1-ohm shunt resistance of the lower R_2 circuit is B^2, the second-order power of B.
3. Divide B^2 thus obtained by B^{-1} in a similar way using the R_3 circuits. The division is B^3, the third-order power of B, and is across the 1-ohm shunt resistance of the lower R_3 circuit.
4. Similarly, B^n, the nth-order power of B, is finally obtained by dividing B^{n-1} by B^{-1} using the R_n circuits. The nth-order power of n is across the 1-ohm shunt resistance of the R_n lower circuit when all the n ammeters of the system read unity current.

Again the accuracy of this type of circuit is totally dependent on the relative magnitudes of the different resistances used. The accuracy may be a maximum when, with any successive two values of resistances, R_{n-1} is neglected relative to R_n.

Positive Nth-Order Root

To obtain B if B^n is known, where n is a positive integer, Fig. 5 may be used in the following way:

1. Apply the 1 volt where indicated.
2. Apply a variable voltage at B terminals.
3. Measure B^n at its proper place under the correct operation conditions, i.e., with all ammeters reading unity current for each value of B, until finally the known value B^n appears at its proper place. Under these conditions the required nth root of B^n is equal to B, the voltage applied at the input terminals of the lower circuit of Fig. 5.

Paper 57-746, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted February 21, 1956; made available for printing May 15, 1957.

M. S. M. Abou-Hussein is with the University of Cairo, Giza, Egypt.

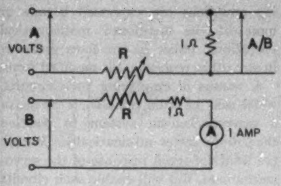


Fig. 1. Division

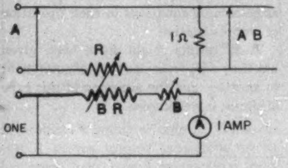


Fig. 2. Multiplication

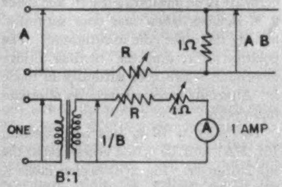


Fig. 3. Multiplication; experimental representation

NEGATIVE NTH-ORDER ROOT

To obtain B^{-1} if B^n is known, where n is a positive integer, Fig. 5 may be also used as follows.

Repeat the three steps mentioned in connection with the case of the nth-order positive root. When B is finally obtained, the required negative nth-order root B^{-1} may be directly measured where indicated across the 1-ohm shunt resistance of the upper circuit of Fig. 5.

The negative nth-order root B of the quantity B^{-n} may be also obtained by using Fig. 4. Apply the 1 volt to the unity-volt circuits and vary the voltage B applied to the lowest circuit, in steps, adjusting for each step the different variable resistances to allow the 1 ampere to flow through the different ammeters. Measure, for each step, the value B^{-n} at its proper place until it becomes finally equal to the assumed known value of this quantity. Under these conditions the applied voltage B is the required negative nth root of B^{-n}.

LOGARITHM

$\operatorname{Log}_B x = n$ is considered, where B and n are known and n is a positive or negative integer. If n is assumed to be a positive integer Fig. 5 may be used. The B- and 1-volt supplies are to be applied where indicated and the previously mentioned adjustments repeated. When all ammeters read unity, one of the values B^2, B^3, $B^4, \ldots$, B^n, say B^y, will be found equal to the known value x. Under these conditions the required unknown value n is equal to y.

If, however, n is assumed to be a negative integer, Fig. 4 may be used. The B- and 1-volt supplies are switched on where indicated and the necessary adjustments made. When all the ammeters read unity current it will be found that one of the values B^{-1}, $B^{-2}, \ldots, B^{-n}$ will be equal to the known value x; if this value is B^{-y}, then the required unknown value n is equal to y.

ANTILOGARITHM

$\operatorname{Log}_B x = n$ is considered, where B and n are known and n is a positive or negative integer. Since

$$x = B^n \tag{6}$$

and B and n are known, then the problem goes back to the case of the nth-order power if n is a positive integer, and to the case of the nth-order inversion if n is assumed to be a negative integer. Hence the unknown value x may be found as previously mentioned for these cases.

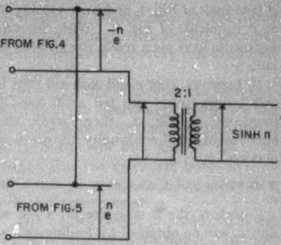


Fig. 6. Hyperbolic sine

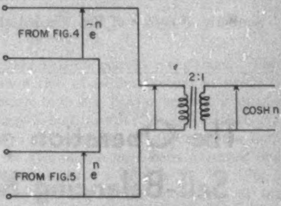


Fig. 7. Hyperbolic cosine

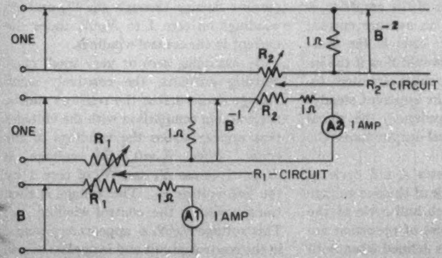
HYPERBOLIC FUNCTIONS OF ANGLE N

The hyperbolic functions of angle n are considered where n is a positive integer. If, in Figs. 4 and 5, B is replaced by the exponent e, then values of $e^{\pm n}$ may be easily obtained, as previously mentioned, with the negative value of n from Fig. 4 and positive value of n from Fig. 5. Since

$$\left.\begin{aligned}\sinh n &= (e^n - e^{-n})/2\\ \cosh n &= (e^n + e^{-n})/2\end{aligned}\right\} \tag{7}$$

then Figs. 6 and 7 may give respectively the sinh and cosh functions.

Tanh n is the divison of sinh n (equal to A for example) and cosh n (equal to B for example). This division (A/B) may be obtained as mentioned in connection with Fig. 1 by feeding sinh n from Fig. 6 at



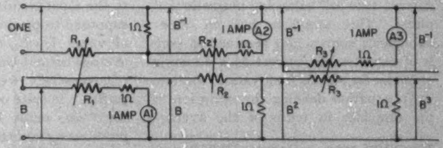


Fig. 4 (left). Inversion, finding roots, logarithms, and antilogarithms

Fig. 5 (above). Finding powers, roots, logarithms, and antilogarithms

A terminals of Fig. 1, and cosh n of Fig. 7 at *B* terminals of Fig. 1. When the adjustments have been made the division is tanh n.

If, however, sinh n and cosh n of Figs. 6 and 7 respectively are fed at *B* and *A* terminals of Fig. 1, then, when the ammeter of Fig. 1 reads unity, the division obtained is

$$A/B=\cosh n/\sinh n=\coth n \tag{8}$$

Cosech n, the first-degree inversion of sinh n, can be obtained by feeding sinh n of Fig. 6 at *B* terminals of Fig. 4, and the first-degree inversion is

$$1/B=1/\sinh n=\operatorname{cosech} n \tag{9}$$

Similarly, if cosh n of Fig. 7 is fed at *B* terminals of Fig. 4, then the first-degree inversion is

$$1/B=1/\cosh n=\operatorname{sech} n \tag{10}$$

Hence all the hyperbolic functions of the angle n may be obtained as mentioned using linear simplified circuits.

Conclusions

It is thus possible to perform many nonlinear mathematical computations by linear elements in a simple systematic way and by use of the ordinary measuring instruments, ammeters and voltmeters.

The importance of developing the given circuits is in solving easily and quickly many physical phenomena whose governing equations are expressed in one or more of the mentioned mathematical operations, either in the form given or in any other representable modified form.

A means of employing such circuits, which may be often valuable, is the use of a servomechanism system to position the potentiometer automatically. Within the limits of speed response of the servomechanism, this will enable such circuits to be fed to the different analog computers having the given mathematical operations as governing equations in their operation or representation.

Tests results might have been given but they vary from one type of equipment to another, and the accuracy of this type of circuit is well known.

The Operation of the Current-Type Self-Balancing Magnetic Amplifier

ALBERT D. KRALL
NONMEMBER AIEE

SELF-BALANCING magnetic amplifiers[1,2] employ feedback to produce extremely useful characteristics. The input impedance can be made to approach either zero or infinity, the transfer characteristic is linear, the response time is easily varied to the minimum delay of one-half cycle, and the drift can be low because of a high signal-to-noise ratio. Applications of these circuits are appearing in increasing numbers in the literature[3–6] but until recently an analytical study was not available because of the complexity of the feedbacks involved in their operation. A previous paper[6] indicates the general principles of self-balancing operation and discusses in detail one type of self-balancing circuit, the voltage-type. The following discussion will pertain to the current-type self-balancing magnetic amplifier. This amplifier has an input impedance approaching zero and as such is of considerable interest as a summing amplifier for signals from various sources.

An equation defining the operation of the amplifier in terms of the average input and output voltage is derived. This equation contains feedback dependent on the circuit parameters, which can be made negative or positive. The resolution of this feedback term into composite magnetic and electric feedbacks[7] is left to those readers well versed in the prior art.

Circuit Operation

Fig. 1 is a schematic of the circuit to be described. This amplifier operates properly with the polarity of the control voltage *Ec*, as shown on the diagram. It is possible to construct reversible polarity input circuits. The signal source is considered to be direct voltage *Ec* plus the series resistance *Rc* and, when connected to the input terminals of the amplifier, is supposed to produce an average current of value $I_c=E_c/R_c$. This is the same average current that would flow if the input terminals were short-circuited together. In spite of the apparent steady-state zero input impedance, the signal produces a proportional output across the load resistor R_L.

The circuit completes a full cycle of operation in a full cycle of the line voltage e_s frequency. In each half-cycle of the line voltage, two modes of operation are permitted. Mode 1 is defined when both cores are in an unsaturated state and mode 2 is defined when one core saturates, leaving the other core unsaturated. The polarity of e_s and the rectifier bridge determine which core saturates in mode 2. Alternate cores saturate on alternate half-cycles.

The polarity of e_s has been chosen, Fig. 2(A), and the period is designated the nth half-cycle. Fig. 2(A) depicts mode 1, i.e., cores 1 and 2 are unsaturated. Perfect rectifiers are assumed, and to simplify the drawing are omitted, but it must be remembered that the rectifiers will switch the connections to the windings on alternate half-cycles. One further simplification was made in Fig. 2(A). Notice the N_f and N_s windings on core 2 are series opposing or bifilar for the nth half-cycle. The number of turns on N_f and N_s are restricted to be equal. Therefore the magnetic effects of these windings are opposing and need not be considered. Fig. 2(B) contains the essential necessary to explain mode 1 of the nth half-cycle. With core 2 unsaturated, the current through the control windings N_c is limited to the magnetizing current of the control winding on core 2. This in turn limits the current flowing through the N_s and N_f windings on core 1 to N_c/N_s times the current in the control windings.

By assuming zero or very small magnetizing currents, the relatively small voltage drops across the resistors can be neglected in comparison with the voltages that appear across the windings of the cores. When $E_c=0$, the voltage is dropped across N_s and N_f of core 1 by the line voltage e_s. The voltage is then transformed to the control winding N_{c_1}. This voltage $N_c/N_s\,e_1$ appears as a source in the control circuit and is applied across

---

Paper 57-765, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted September 17, 1956; made available for printing May 3, 1957.

ALBERT D. KRALL is with the Naval Ordnance Laboratory, Silver Spring, Md.

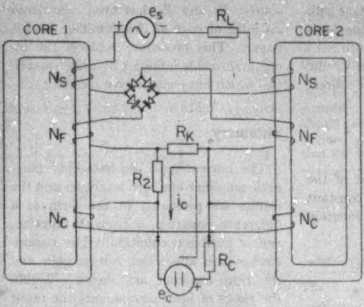


Fig. 1. Current-type self-balancing magnetic amplifier

the control winding of core 2. Thus, as core 1 moves toward saturation on its hysteresis loop, core 2 moves away from saturation an equal amount on its hysteresis loop. The cores are designed with the ability to absorb the voltage e_1 on the N_s winding without saturating. Therefore, if $Ec=0$, mode 1 will continue for the entire nth half-cycle and in the $(n+1)$th half-cycle the rectifier bridge will switch the roles played by each core. Now, when $E_c \neq 0$, a simultaneous solution of the loop equations reveals that the voltage e_1 increases from the quiescent condition while the voltage across the control winding of core 2 decreases. The increase in e_1 causes core 1 to saturate and the decrease in voltage across core 2 means that less flux was reset in core 2 during mode 1. Since core 1 does saturate before the half-cycle ends, mode 2 must be considered to see its effect on the flux reset in core 2 and to see how the circuit produces an output. Fig. 3(B) is the schematic representation of mode 2 of the nth half-cycle. Core 1 has been omitted because it is now saturated and the resistor R_w which contains all the winding and rectifier resistances in the load circuit, has been inserted.

The current through N_{c2} is still limited to magnetizing current, but i_c and i_L are now allowed to increase. The voltage drop across R_2 and R_k is applied to Nc_2 and continues the process of resetting flux in core 2. Now, if any two successive half-cycles produce equal reset fluxes in the cores, even though the control voltage changes, all succeeding half-cycles will do likewise and steady-state conditions will exist. In other words, the time constant[8] of the amplifier would be zero, leaving a dead-time[8] or a transportation lag of less than one half-cycle, i.e., output occurs on the same half-cycle that input is applied. In an effort to achieve this condition the average voltage across core 2 (the reset voltage) was summed for quiescent conditions and summed for signal conditions. No adjustment in parameters could make these voltages equal and thus the time constant could not be made zero. An equilibrium can be reached in the second half-cycle by a suitable adjustment of parameters shown in equation 31.

In Fig. 3(B) it can be seen that the current i_L is limited only by the value of the line voltage and a resistor network. Output power is therefore delivered during mode 2.

So far the nth half-cycle has been considered. The equivalent circuits for the $(n+1)$th half-cycle (i.e., the polarity of e_s reversed) will look exactly the same except that the cores will have reversed their operation with core 1 now resetting flux and core 2 saturating.

The control circuit was said to present a zero impedance to the signal. In other words the current iC should have the value of E_c/R_c. A glance at Fig. 2(A) shows that

$$i_c=\frac{E_c}{R_c+R_2+R_k}$$

if magnetizing current is neglected. This reduction of current occurs in mode 1, while from Fig. 3(A) or mode 2 it can be shown that current iC is greater than E_c/R_c. In Appendix II, the currents from the two modes have been averaged and found to have a value of E_c/R_c. Fig. 4 is a drawing of the waveshape of the current iC. The impedance is certainly not zero instantaneously but it does appear to be zero on the average. A capacitor across the input terminals will filter a large percentage of the alternating current without affecting the operation adversely.

Interpretation of Analytical Results

In Appendix I the equation defining the operation of the amplifier is written

$$E_o(n+1)=AE_c(n)+BE_c(n+1)+CE_o(n)$$

where the coefficients are given in equation 21. This equation states that the average output voltage on a given half-cycle has three components. The components are the control voltage on the

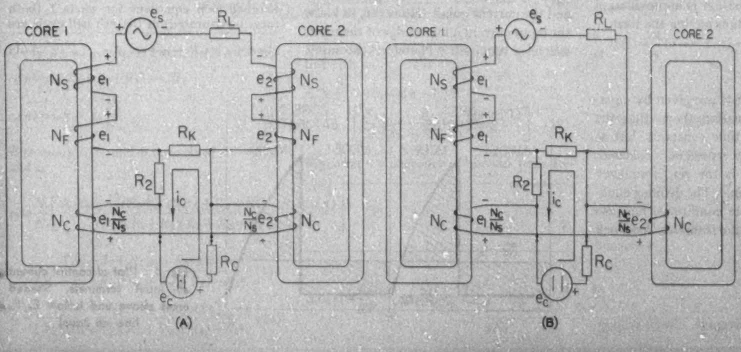


Fig. 2. A—Equivalent circuit of Fig. 1 for the nth half-cycle, mode 1. B — Equivalent circuit of A

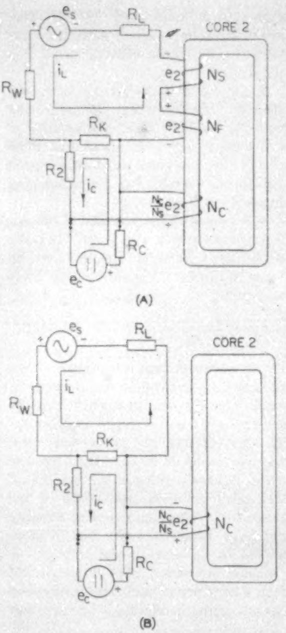


Fig. 3. A—Equivalent circuit of Fig. 1 for nth half-cycle, mode 2. B—Equivalent circuit of A

previous half-cycle, the control voltage on the present half-cycle, and the output voltage on the previous half-cycle. When the coefficient $C=0$, the output is dependent only on the control voltage or, in other words, there is no feedback present. Since the output depends on the control of the previous half-cycle the minimum response time has a delay of one-half-cycle.

The defining equation is approximated by a transfer function having the form

$$KG(s)=K\frac{\tau_1 s+1}{\tau_2 s+1}$$

where the coefficients are given by equation 29. This equation shows that the amplifier has two time constants, but as long as $\tau_2 >> \tau_1$ only τ_2 need be considered. The coefficient K is the zero frequency average voltage gain. The defining equation, in steady-state conditions, has also been written in the familiar feedback form

$$\frac{E_o}{E_c}=K=\frac{K_f}{1-\beta K_f}$$

Where K is the average direct voltage gain with feedback, K_f is the voltage gain without feedback, and βK_f is the feedback factor.

From any of the three forms of the defining equation the average direct voltage gain can be found as

$$K=\frac{R_L(R_2+R_k)}{R_cR_k}$$

The voltage gain is independent of the turns on the cores. The time constant of the amplifier from the transfer function is

$$\tau_2=\frac{N_c}{2N_s}\frac{(R_c+R_2+R_k)(R_L+R_w+R_k)-R_k^2}{R_cR_k}\frac{\pi}{\omega}$$

The time constant is dependent on the ratio of turns on the core. Therefore the voltage gain and the time constant can be set independently. The consequence of this is that the figure of merit K/τ_2, is variable within limits.

General Comments

The previous literature usually associates an infinite power gain with the self-balancing amplifiers. In this paper equation 37 shows the power gain to be finite. This apparent discrepancy is due merely to the definition of the input terminals. If the input terminals of the amplifier (i.e., excluding the signal e_c and its resistance R_c) are considered, it is easy to show that the amplifier has an infinite voltage and power gain. The input terminals are shown to have zero impedance in Appendix II. Thus no average voltage or power appears at the input terminals and the average voltage and power gains must then be infinite. Since a zero input power and an infinite power gain cannot be used to calculate the output power, it appears desirable to take a different point of view where the gains are all finite. It is possible to keep the infinite voltage and power gain concepts and completely describe the amplifier with the input current and the current gain. However, to know the input current a knowledge of the input source is required. Therefore the input source, Ec and Rc, has been considered an integral part of the amplifier in this paper. This procedure replaces the incomprehensible infinities by finite quantities which have qualitative meaning.

Summary

The current-type self-balancing magnetic amplifier has been analyzed and the results are presented in the form of a difference equation, a transfer function, and a feedback equation. The parameters which affect the voltage gain and the time constant are shown. Within the range of ideal components the input impedance is shown to be zero.

Appendix I. Analysis of Current-Type Self-Balancing Magnetic Amplifier

The basis[9] for this analysis is Faraday's law, which states that an average voltage appearing across a winding changes the flux in the core material. The reactors in the current-type self-balancing magnetic amplifier operate from a given point B_r on the hysteresis loop. During any given cycle of the carrier frequency they are considered to start from B_r on the hysteresis loop, travel a flux path, and return to B_r at the end of the complete cycle. Because they return to the starting point B_r, it can be said that the net change in flux over the complete cycle is zero. Furthermore, Faraday's law would then indicate that there is no net average voltage across the windings during that period.

Fig. 2(A) is a representation of Fig. 1 during the *n*th half-cycle, mode 1. The rectifiers have been assumed ideal, the polarity of supply for the *n*th half-cycle is indicated, the voltages e_1 and e_2 have been assigned to the N_s windings of core 1 and core 2 respectively, and the magnetizing currents which flow through the windings have been neglected. Throughout the analysis the windings N_s and N_f are equal, and thus the voltages across N_s and N_f on a given core are equal. From Fig. 2 the Kirchoff loop equations for mode 1 (both cores unsaturated) of the *n*th half-cycle are

$$e_s=e_1+e_1-i_cR_k-e_2+e_2 \qquad (1)$$

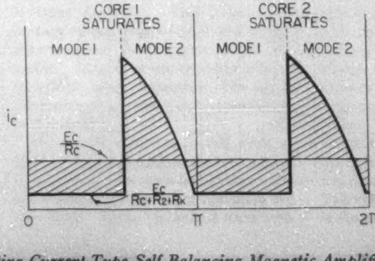


Fig. 4. Plot of control current at input terminals. Shaded areas above and below E_c/R_c line are equal

$$e_1\frac{N_c}{N_s}=e_2\frac{N_c}{N_s}+i_c(R_2+R_k) \qquad (2)$$

$$e_1\frac{N_c}{N_s}=e_2\frac{N_c}{N_s}-i_cR_c+e_c(n) \qquad (3)$$

Equations 1, 2 and 3 are solved for e_1, e_2, and i_c

$$e_1=\frac{1}{2}\left[e_s+\frac{R_k}{R_c+R_2+R_k}e_c(n)\right] \qquad (4)$$

$$e_2=\frac{1}{2}\left[e_s-\frac{\frac{2N_s}{N_c}(R_2+R_k)-R_k}{R_c+R_2+R_k}e_c(n)\right] \qquad (5)$$

$$i_c=\frac{e_c(n)}{R_c+R_2+R_k} \qquad (6)$$

Considering $ec=0$, equations 4 and 5 show that the voltage across the N_s winding of cores 1 and 2 is one half the line voltage e_s on each. The cores must have the ability to absorb at least this much voltage. Also, it may be seen from equations 4 and 5 that when $e_c\neq 0$, e_1 is greater than e_2. The increase in the voltage e_1, when $e_c\neq 0$, causes core 1 to saturate, leaving core 2 unsaturated. Equation 6 gives the value of signal current for mode 1 and shows that it is reduced from the value e_c/R_c. Thus the input terminals do not exhibit zero impedance during mode 1.

Consider now that core 1 has saturated and core 2 has remained unsaturated as indicated by equations 4 and 5. This condition will define mode 2 of the nth half-cycle and figure (3A) is the equivalent circuit for mode 2. The load current i_L is no longer magnetizing current as it was in mode 1, since the N_s and N_F windings on core 2 are opposing. The resistor R_w has been inserted in the load circuit to account for the now appreciable voltage drop across the forward resistance of the rectifiers, the winding resistance, and the power source resistance. The current through the control winding N_c is still limited to magnetizing current and is neglected.

The Kirchoff loop equations for figure 3(A) are equations 7, 8, and 9.

$$e_s=i_L(R_L+R_w+R_k)-i_cR_k+e_1-e_2 \qquad (7)$$

$$e_c(n)=i_c(R_c+R_2+R_k)-i_LR_k \qquad (8)$$

$$e_c(n)=i_cR_c-e_2\frac{N_c}{N_s} \qquad (9)$$

Equations 7, 8, and 9 are solved for e_2, i_L, and i_c

$$e_2=\frac{N_s}{N_c}\left[\frac{R_cR_ke_s-[(R_2+R_k)(R_L+R_w+R_k)-R_k^2]e_c(n)}{(R_c+R_2+R_k)(R_L+R_w+R_k)-R_k^2}\right] \qquad (10)$$

$$i_L=\frac{(R_c+R_2+R_k)e_s+R_ke_c(n)}{(R_c+R_2+R_k)(R_L+R_w+R_k)-R_k^2} \qquad (11)$$

$$i_c=\frac{R_ke_s+(R_L+R_w+R_k)e_c(n)}{(R_c+R_2+R_k)(R_L+R_w+R_k)-R_k^2} \qquad (12)$$

The voltage e_2 in equation 10 is not allowed to be enough to saturate core 2 by the end of the nth half-cycle, and so there are just two modes of operation in a given half-cycle. To complete a full cycle of operation and thus be able to average the voltage over a complete cycle, the $(n+1)$th half-cycle should now be considered. However, the symmetry of the circuitry in Fig. 1 reveals that the voltages on core 1 in the nth half-cycle are the voltages that appear on core 2 in the $(n+1)$th half-cycle, etc., i.e.

$$e_1(n)=e_2(n+1) \qquad (13)$$

$$e_2(n)=e_1(n+1) \qquad (14)$$

$$i_c(n)=i_c(n+1) \qquad (15)$$

$$i_L(n)=i_L(n+1) \qquad (16)$$

Before these voltages may be averaged and added for a complete cycle, a relationship of the relative duration of modes 1 and 2 must be written. In the nth half-cycle core 1 saturates and defines the end of mode 1 and the beginning of mode 2. Power is delivered to the load during mode 2 only. If the area of the supply voltage half-cycle is considered to be 1 (unity) and if $a(n)$ is defined as the fraction of the total area coinciding with mode 2, then the fraction of the total area coinciding with mode 1 is $1-a(n)$. Fig. 5 is a pictorial representation of $a(n)$ and similarly $a(n+1)$. The average output voltage in the nth half-cycle defined in terms of $a(n)$ will be

$$E_0(n)=a(n)i_L(n)R_L \qquad (17)$$

Enough information has now been found to write the average voltage across core 2 for the complete cycle of operation. The average voltage across core 2 for the nth half-cycle is

$$E_2(n)=[1-a(n)]\left[\frac{E_s}{2}-\frac{\frac{2N_s}{N_c}(R_2+R_k)-R_k}{R_c+R_2+R_k}\frac{E_c(n)}{2}\right]+ a(n)\frac{N_s}{N_c}\frac{R_cR_kE_s-[(R_2+R_k)(R_L+R_w+R_k)-R_k^2]E_c(n)}{(R_c+R_2+R_k)(R_L+R_w+R_k)-R_k^2} \qquad (18)$$

The average voltage across core 2 for the $(n+1)$th half-cycle is

$$E_2(n+1)=[1-a(n+1)]\times\left[\frac{E_s}{2}+\frac{R_k}{R_c+R_2+R_k}\frac{E_c(n+1)}{2}\right] \qquad (19)$$

Equation 19 contains no second term for mode 2 in the $(n+1)$th half-cycle because core 2 is saturated, thus no voltage appears across the winding. As was explained earlier the average voltage across core 2 for the complete cycle is zero, thus

$$E_2(n)=E_2(n+1) \qquad (20)$$

Combining and reducing equations 11, 17, 18, 19, and 20, and solving for the average output on the $(n+1)$th half-cycle results in

$$E_0(n+1)=\frac{R_L\left[\frac{2N_s}{N_c}(R_2+R_k)-R_k\right]E_c(n)+R_LR_kE_c(n+1)}{(R_c+R_2+R_k)(R_L+R_w+R_k)-R_k^2}+ \left[1-\frac{2N_s}{N_c}\frac{R_cR_k}{(R_c+R_2+R_k)(R_L+R_w+R_k)-R_k^2}\right]E_0(n) \qquad (21)$$

In reducing equation 21 the control voltage E_c has been assumed small in comparison with the line voltage E_s. Equation 21 defines the operation of the amplifier. For steady-state conditions $E_0(n+1)=E_0(n)$ and $E_c(n+1)=E_c(n)$. Substituting the steady-state conditions yields

$$E_0=\frac{R_L(R_2+R_k)}{R_cR_k}E_c \qquad (22)$$

The coefficient of the control voltage E_c is the steady-state average direct voltage gain. It is independent of the turns ratio.

Rewriting equation 21 where the symbols A, B and C are the new coefficients gives

$$E_0(n+1)=AE_c(n)+BE_c(n+1)+CE_0(n) \qquad (23)$$

Writing the equalities gives

$$E_0(n+1)=[(E_0(n+1)-E_0(n)]+E_0(n) \qquad (24)$$

$$E_c(n+1)=[E_c(n+1)-E_c(n)]+E_c(n) \qquad (25)$$

and substituting into equation 23 and dividing through by π ω yields

$$\frac{E_0(n+1)-E_0(n)}{\frac{\pi}{\omega}}+\frac{(1-c)E_0(n)}{\frac{\pi}{\omega}}= B\frac{E_c(n+1)-E_c(n)}{\frac{\pi}{\omega}}+(A+B)\frac{E_c(n)}{\frac{\pi}{\omega}} \qquad (26)$$

Equation 26 may be approximated by a differential equation when the frequency of E_s is high compared to E_c. The differential equation is

$$\frac{d}{dt}E_0(t)-\frac{\omega}{\pi}(1-c)E_0(t)=B\frac{d}{dt}E_c(t)+\frac{\omega}{\pi}(A+B)E_c(t) \qquad (27)$$

The Laplace transform of equation 27 is

$$s\mathcal{L}E_0(t)+\frac{\omega}{\pi}(1-c)\mathcal{L}E_0(t)=B_s\mathcal{L}E_c(t)+\frac{\omega}{\pi}(A+B)\mathcal{L}E_c(t) \qquad (28)$$

From equation 28 the approximate transfer function is

$$KG(s)=\frac{\mathcal{L}E_0(t)}{\mathcal{L}E_c(t)}=\frac{R_L(R_2+R_k)}{R_cR_k}\,\frac{\frac{N_c}{2N_s}\frac{R_k}{R_2+R_k}\frac{\pi}{\omega}s+1}{\frac{N_c}{2N_s}\frac{(R_c+R_2+R_k)(R_L+R_w+R_k)-R_k^2}{R_cR_k}\frac{\pi}{\omega}s+1} \qquad (29)$$

The transfer function is of the form

$$KG(s)=K\frac{\tau_1 s+1}{\tau_2 s+1} \quad (30)$$

If $\tau_1 << \tau_2$, then T_2 is the time constant of the amplifier. To make the amplifier respond to one-half cycle of the supply or carrier frequency let $\tau_2=\pi/\omega$ or

$$\frac{N_c}{2N_s}\frac{(R_c+R_s+R_k)(R_L+R_w+R_k)-R_k^2}{R_cR_k}=1 \quad (31)$$

Appendix II. Zero Input Resistance

One of the main purposes of this amplifier is to provide a zero input resistance, i.e., the average current flowing in the control circuit should have a value of E_c/R_c. In the following, the average currents from mode 1 and mode 2 flow in the input circuit will be summed to show that this is true. Fig. 4 is a typical plot of the current. From equation 6 and the factor $a(n)$ in Appendix I the average input current for mode 1 is

$$I_c=[1-a(n)]\frac{E_c}{R_c+R_s+R_k} \quad (32)$$

From equation 12 and the factor $a(n)$ in Appendix I the average input current for mode 2 is

$$I_c=a(n)\,\frac{R_kE_s+(R_L+R_w+R_k)E_c}{(R_c+R_s+R_k)(R_L+R_w+R_k)-R_k^2} \quad (33)$$

Substituting for the value of $a(n)$ and adding equations 32 and 33 yields

$$I_c(\text{mode 1})+I_c(\text{mode 2})=\frac{E_c}{R_c} \quad (34)$$

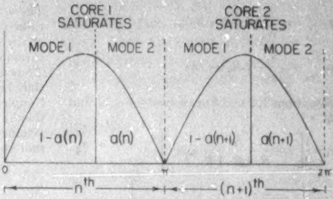
Equation 34 indicates that the average input current is the same as if the amplifiers terminals were short-circuited. Since the average input current is E_c/R_c, the average input power is

$$P_{in}=E_cI_c=\frac{E_c^2}{R_c} \quad (35)$$

From equation 22 the average output voltage is $E_0=\frac{R_L(R_s+R_k)}{R_cR_k}E_c$ and thus the average output power is equation 36, following

$$P_{out}=\frac{E_0^2}{R_L}=\frac{R_L(R_s+R_k)^2}{R_c^2R_k^2}E_c^2 \quad (36)$$

Fig. 5. Pictorial definitions of a(n) and a(n+1)



Combining equations 35 and 36 yields the average power gain

$$P_{gain}=\frac{P_{out}}{P_{in}}=\frac{R_L(R_s+R_k)^2}{R_cR_k^2} \quad (37)$$

The figure of merit can be defined as the voltage-gain divided by the time constant. The voltage gain is given by equation 29, when $s=0$ and the time constant given by equation 29 and 30.

$$\frac{K}{\tau_2}=\frac{2N_s}{N_c}\left[\frac{R_L(R_s+R_k)}{(R_c+R_s+R_k)(R_L+R_w+R_k)-R_k^2}\right]2\text{ frequency} \quad (38)$$

As Johannessen points out, the figure of merit defined this way is synonymous with the gain-bandwidth product of an electronic amplifier. The gain-bandwidth product is known to be independent of feedback and is uniquely determined, as is equation 38, by the forward gain of the amplifier. The forward gain of the current-type amplifier is seen from equation 21 by summing the coefficients of $E_c(n)$ and $E_c(n+1)$ while eliminating the feedback, the coefficient of $E_0(n)$. Thus equation 38 can be written

$$\frac{K}{\tau_2}=[\text{forward gain}]2\text{ frequency} \quad (39)$$

From the equation $K=K_f/(1-\beta K_f)$ the feedback factor of the amplifier can be found where

$K=$ gain with feedback $=\frac{R_L(R_s+R_k)}{R_cR_k}$

$K_f=$ gain without feedback

$$=\frac{2N_s}{N_c}\frac{R_L(R_s+R_k)}{(R_c+R_s+R_k)(R_L+R_w+R_k)-R_k^2}$$

$\beta=$ fraction of the output applied to the input

In this equation β is positive when feedback aids the input signal. Substituting and solving for β

$$\beta=\frac{(R_s+R_c+R_k)(R_L+R_w+R_k)-R_k^2-\frac{2N_s}{N_c}R_cR_k}{\frac{2N_s}{N_c}(R_s+R_k)R_L} \quad (40)$$

When the numerator is zero the feedback is zero. This is the identical condition required for half-cycle response or for $\tau_2=$ one half-cycle. The feedback factor βK_f is the ratio of the feedback voltage to the applied control voltage

$$\beta K_f=1-\frac{2N_s}{N_c}\frac{R_cR_k}{(R_c+R_s+R_k)(R_L+R_w+R_k)-R_k^2} \quad (41)$$

The feedback factor can also be seen in the defining equation 21, as the coefficient of the $E_0(n)$ term. The defining equation may now be written for steady state as

$$E_0=K_fE_c+\beta K_fE_0 \quad (42)$$

Or upon rearranging terms

$$\frac{E_0}{E_c}=K=\frac{K_f}{1-\beta K_f} \quad (43)$$

References

1. MAGNETIC AMPLIFIERS OF THE SELF-BALANCING POTENTIOMETER TYPE, W. A. Geyger. *AIEE Transactions*, vol. 71, pt. I, Jan. 1952, pp. 383–95.

2. SELF-BALANCING MAGNETIC AMPLIFIER, W. A. Geyger. *U. S. Patent no. 2,700,130*, Jan. 18, 1955.

3. AN ANALOGUE COMPUTER TECHNIQUE USING MAGNETIC AMPLIFIERS, B. R. Davis, I. H. Swift. *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 635–40.

4. AN OPERATIONAL-TYPE MAGNETIC AMPLIFIER FOR AIR-BORNE SERVO-CONTROL SYSTEMS, R. M. Hubbard. *Ibid.*, Nov. 1954, pp. 425–30.

5. 60-CYCLE SELF-BALANCING MAGNETIC SERVO AMPLIFIER, W. A. Geyger. *AIEE Special Publication T-86*, "Conference on Magnetic Amplifiers," Apr. 1956, pp. 221–31.

6. THE OPERATION OF THE SELF-BALANCING MAGNETIC AMPLIFIER, Albert D. Krall, Edward T. Hooper. *AIEE Transactions*, vol. 75, pt. I, Mar. 1956, pp. 79–84.

7. SELF-BALANCING MAGNETIC AMPLIFIERS OF THE DIFFERENTIAL FEEDBACK TYPE, W. A. Geyger. *Ibid.*, vol. 76, pt. I, Mar. 1957, pp. 39–46.

8. MAGNETIC AMPLIFIERS (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 146–50.

9. ANALYSIS OF MAGNETIC AMPLIFIERS BY THE USE OF DIFFERENCE EQUATIONS, P. R. Johannessen. *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 700–11.

A New 5-Channel Carrier System for Subscriber Use Employing Transistors

B. R. STACHIEWICZ
MEMBER AIEE

R. L. LAYBURN
NONMEMBER AIEE

D. F. JAMIESON
NONMEMBER AIEE

THE NEW Subscriber-Carrier System type *561* described in this paper is a result of many months of engineering effort by many individuals in the Stromberg-Carlson organization. The design was undertaken to satisfy a pressing need in the telephone industry for an inexpensive yet good quality system which would help alleviate the critical situation existing in the field of rural service.

An attempt was made to take advantage of the most modern techniques in both the mechanical and electrical design, and to combine the low cost of the equipment with high reliability and good transmission quality. The necessity of maintenance has been reduced to the minimum and the initial line-up procedure has been simplified.

One of the first decisions that had to be made concerned the use of transistors in place of tubes. A study of commercially available units indicated that a high degree of product uniformity has been achieved by several manufacturers and consequently the variations in transistor parameters can be held to close tolerances. Further tests indicated that with proper attention to circuitry details, reliable operation of germanium units could be obtained at temperatures up to 150 degrees Fahrenheit.

Another attractive feature leading to the substitution of transistors for tubes was the small size, low power drain, and theoretically infinite life of transistors. It was, therefore, decided to adopt them as a basic circuit element in the new carrier system.

In the course of development many problems were encountered in the application of these new units. Special solutions had to be found and lengthy investigations were necessary, before satisfactory performance was obtained. Most of these problems were connected with the operation of the circuits at higher temperatures and with voltage surges in circuits where pulses of current were used for signaling. Here the requirements for high collector current conflicted with the necessity of not exceeding the dissipation rating of available transistors.

Applications

The system is designed to permit the use of five carrier channels over the available rural lines without the necessity of any modification of the existing central office equipment.

The frequency allocation is shown in Fig. 1. The choice of carrier frequencies was dictated by several considerations. The lowest frequency was chosen to cause no interference with voice circuits. The highest frequency was kept below 140 kc since the study of available wire facilities indicated that line attenuation and crosstalk problems at higher frequencies would be objectionable. Intermediate frequencies were chosen with a view to facilitating the design of the filters.

Considerations of near- and far-end crosstalk conditions in the terminal equipment, with five channels in parallel, dictated the choice of transmitting and receiving allocations.

Double side-band transmission was chosen in preference to other methods, mainly because of the less expensive filters and the elimination of carrier oscillators for demodulation.

The signaling circuitry is designed to handle subscriber dialing and either bridged or divided ringing from the central office. In this way, facilities for full-selective multifrequency ringing can be obtained without modification of existing central office equipment. Up to ten party lines per channel can thus be served, increasing the capacity of a single rural line by 50 subscribers.

One of the outstanding features of the system is the flexibility with which equipment can be arranged to suit particular conditions. The system provides five self-contained duplex speech channels which can be added or removed without affecting the operation of remaining units. All channels are identical except for frequency determining networks, and any unit can be converted to operate at any desired frequency by simply plugging in proper filters and oscillators.

The equipment operates from −48 volts d-c or 117 volts a-c and the power supply is designed to accommodate a-c line-voltage variations of ±10%. Provision is made to remove the power supply unit when local battery is available at the central office. An additional power unit is provided at the subscriber end. This extra supply is required to furnish the ringing and "talking battery" power.

Plug-in compandors are provided on an optional basis, and can be added when required without the necessity of rewiring or otherwise disturbing the system. Situations may arise where satisfactory operation may be obtained without compandors; in such cases the removal of these units will result in the reduction of the cost of the terminal. The use of the compandor will depend on noise and crosstalk contributed by the wire line only. Terminals are engineered for operation without compandors, and the noise and crosstalk originating in the carrier equipment is not objectionable when compandors are omitted.

Operating Levels

Each channel is arranged to provide a 4-db (decibel) net loss between 2-wire drops. The nominal voice-frequency (v-f) transmitting level is 0 dbm (db referred to 1 milliwatt) and consequently a nominal v-f output of −4 db is obtained at the far end. The maximum transmitting line level of the carrier is +5 dbm and each side band is transmitted at −8 dbm (referred to 0 level at 2-wire input) or 13 db below carrier.

Any channel will operate satisfactorily with total line attenuation of 35 db between the terminals. When the losses at channel carrier frequency exceed this limit a degradation of transmission should be expected. Nominal input level to the receiver has been set at −18 db ±1 db.

A strap pad is provided at the receiver input to adjust this level when line attenuation is less than 20 db under dry weather conditions. An automatic gain control is also provided. The regulation characteristic of the receiver is shown in Fig. 2. Voice-frequency output will stay within 1 db of normal, for variation in line attenuation resulting in receiver input level change of +5 or −10 db about nominal input.

The physical subscriber line can have a loop resistance of up to 1,200 ohms, including the telephone, and a leakage of

---

Paper 56-802, recommended by the AIEE Wire Communications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted April 16, 1956; made available for printing February 7, 1957.

B. R. Stachiewicz, R. L. Layburn, and D. F. Jamieson are with Stromberg-Carlson, a division of General Dynamics Corporation, Rochester, N. Y.

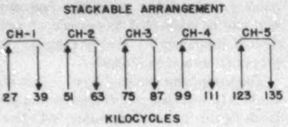


Fig. 1. Frequency allocation chart

15,000 ohms between conductors and ground, for satisfactory ringing.

Equipment

The central office terminal is adapted for standard 19-inch rack mounting and occupies $3^1/_2$ inches of rack space.

The circuitry is divided into subassemblies mounted on printed wire cards. The following cards make up the central office terminal:

1. Terminal card.
2. Receiving filter card.
3. Transmitting filter card.
4. Transmitter card.
5. Receiver card.
6. Central office ringer card.
7. Voice-frequency transmitting card.
8. Voice-frequency receiving card.

When compandors are used, the compressor card replaces the v-f transmitting card and the expandor card replaces the v-f receiving card.

The power supply, relays, hybrid transformer, and attenuation networks are also arranged on a plug-in basis and can easily be removed. All these component parts are mounted in an aluminum housing.

The subscriber terminal consists of a unit identical with the central office terminal plus another panel carrying additional relays and power supply. A third panel equipped with a-c outlets, a line coil, and terminal strip is also provided. These three units are mounted in a weatherproof assembly adaptable for pole mounting.

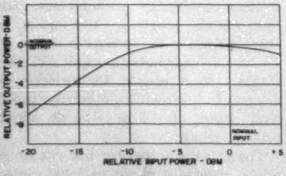


Fig. 2. Receiver regulation characteristic

A central office terminal can be converted into a subscriber terminal by replacing the central office ringer card with the subscriber ringer card, and by the addition of a relay panel. The problem of spare parts is thus greatly simplified.

Printed wiring has been incorporated wherever feasible after a thorough study of advantages and disadvantages in its application was made. Fig. 3 shows a typical plug-in card assembly. Central office and subscriber terminals are shown in Figs. 4 and 5.

Components

The design of filters and other networks used in the system was primarily influenced by the requirements of avilable space and cost. Special design methods were used and the frequency allocation chosen so as to facilitate the filter design. The frequency-attenuation characteristic of the receiving filter is shown in Fig. 6.

Tantalytic capacitors are used where high capacitance with stable low-temperature characteristics is required. The performance of these units proved to be satisfactory and the large capacity combined with small size makes their use in subscriber carrier very attractive. Extensive use is also made of thermistors and germanium diodes. The thermistor properties are utilized in the automatic gain control circuit.

All components were tested for operation under the extremes of temperature before their use in the system was approved.

Circuit Description

The block diagrams of central office terminal and subscriber terminal are shown in Figs. 7 and 8.

Speech signals are transmitted through the v-f hybrid coil to the modulator unit where they are mixed with the carrier supplied by a separate crystal-controlled oscillator. The resulting modulated wave is amplified by a high-gain amplifier consisting of two common emitter stages. Bias stabilization circuits hold transistor dissipation within rated limits at high temperatures.

The carrier output level from the amplifier is approximately +6 dbm. Both series and shunt feedback are used to stabilize the output impedance and provide constant gain under varying conditions of temperature and humidity. The frequency response is flat within 1 db for frequencies up to 140 kc and is shown in Fig. 9.

The filters are connected to the transmission line through a line coil, which is designed to match the terminal to either 600 ohms or a 150-ohm line.

Table I. Operating Characteristics of Type 561 Stromberg-Carlson Subscriber Carrier System

1. **Method of transmission: double side band transmitted carrier**
2. **Channel bandwidth: 250 cycles per second; 2,500 cycles per second at 3-db points**
3. **Transmitting carrier power: +5 dbm nominal**
4. **Transmitting single side-band power: −8 dbm nominal**
5. **Maximum line loss between terminals: 35 db at carrier frequency**
6. **Nominal v-f transmitting power: 0 dbm**
7. **Nominal v-f receiving power: −4 dbm**
8. **Maximum distortion: 5%**
9. **Maximum equipment noise and crosstalk: 20 db adjusted at 0 level**
10. **Maximum d-c loop resistance: 1,200 ohms total**
11. **Maximum leakage between conductors or to ground: 15,000 ohms**
12. **Frequency band for 5-channel system: 24-138 kc**
13. **Operating ambient temperatures: −40 degrees Fahrenheit to +140 degrees Fahrenheit**
14. **Operating voltages**
    - **Subscriber terminal: 115 volts a-c ±10%**
    - **Central office: −48 volts d-c ±10%**
15. **Regulation: output ±1 db for +5-, −10-db variation of input about nominal level**
16. **Power consumption**
    - **Central office: idle, 3 watts; talking, 3.5 watts**
    - **Subscriber terminal (operated from 115 volts a-c): idle, 15 watts; talking, 20 watts average**

The Receiving Circuit

The receiving circuit consists of the receiving filter and a printed-wire plug-in assembly comprising two high-frequency stages of amplification, a detector and an audio power output stage.

Incoming signals are selected by a bandpass filter of special design whose frequency response is shown in Fig. 6. This filter is mounted on a card which can be easily removed from the terminal when necessary. The filter is followed by a resistance network which can be strapped to any attenuation from 0 to 18 db in 3-db steps. Straps are applied as required in the initial line-up to adjust the nominal input level to −18 dbm.

A grounded base amplifier which follows the filter is in the automatic gain control circuit. This configuration was chosen for several reasons, the principal one being the requirement of flat frequency response from 24 to 138 kc. Because of the gain controlling nature of this stage, feedback could not be used and consequently the cut-off frequency depended upon the configuration adopted. A common base arrangement has been employed, since the cut-off frequency is substantially higher than in a common emitter stage.

The grounded base stage is very adaptable for use as a gain regulator due to its high output impedance. Also, the gain

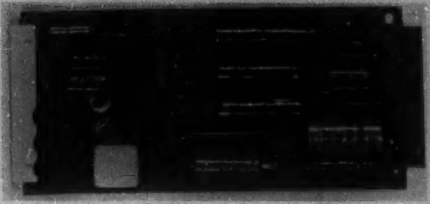


Fig. 3 (above). Typical printed wire card assembly, front

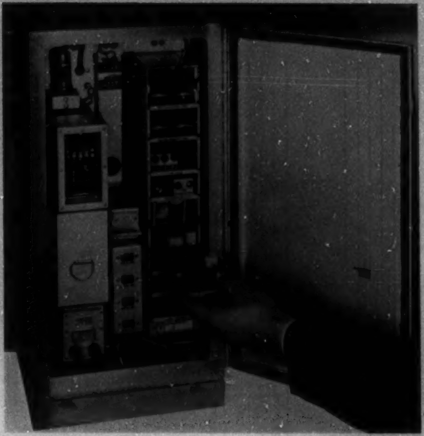


Fig. 5 (right). Subscriber terminal assembly

of such a stage is directly proportional to the load impedance. The use of a thermistor for the load as described in the following affords an easy method for gain control. In addition, the input impedance of the stage is constant over the desired frequency range. This is important from the standpoint of correctly terminating the receiving filter.

The input amplifier is followed by a high-gain feedback-stabilized common emitter amplifier which delivers the signals to the detector-amplifier. The series and shunt feedback in this stage are carefully proportioned as to compensate for decrease in detection efficiency at the high-frequency end of the range.

Carrier frequencies are rectified in the transistor detector-amplifier. The resulting direct current in the collector, which is dependent upon the carrier input and the efficiency of rectification, is passed through two thermistors. One of these thermistors constitutes the a-c load of the grounded base amplifier and the other of the detector-amplifier. The resistance of the thermistors varies in inverse proportion to the rectified carrier current, and consequently output level stabilization is achieved. The variations of the load impedance of the first amplifier and of the detector-amplifier are proportioned so that approximately equal control of gain is achieved in each of these stages. In this way, adequate stiffness ratio is obtained without any back-biasing or d-c amplification. At the same time, overloading of the grounded base stage is effectively prevented over the operating range of input levels. Direct-current amplification has been avoided because of the inherent instability of transistor d-c amplifier circuits.

Initially, a gain control system utilizing the variations of the output impedance of a transistor under the control of base current was tried. However, this circuit proved unsatisfactory because of the necessity of an additional d-c amplifier to provide sufficient current control. Also distortion generated was prohibitive. Use of a separate diode detector was investigated but was discarded in favor of the transistor detector-amplifier, in order to avoid d-c amplification and at the same time achieve economy in components used.

The detector is followed by a 3-kc low-pass filter which eliminates higher frequencies, and delivers signals to the audio power amplifier. A single-transistor common-emitter stage is used. High stability of gain and of the output impedance has been achieved by proper design of the series and shunt feedback networks. Power output of approximately +6.5 dbm was obtained from this amplifier, with relatively low distortion.

The signal relay amplifier in the central office terminal is driven by signals picked off at the input to the detector. Since the relay must be operated when the subscriber terminal transmits carrier, the incoming signal is first amplified in a feedback-stabilized stage and is then rectified in the base circuit of the relay amplifier. The collector current operates the signal relay whenever carrier is on, and consequently the relay follows pulses of carrier current produced by subscriber dialing. The advantage of this type of operation lies in the reduction of current drain, since the carrier transmission from the subscriber terminal is normally off.

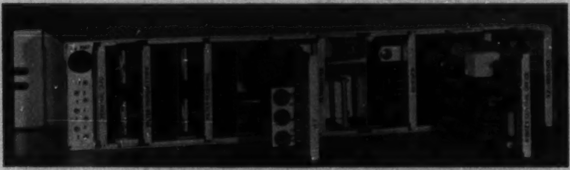


Fig. 4 (above). Central office terminal assembly

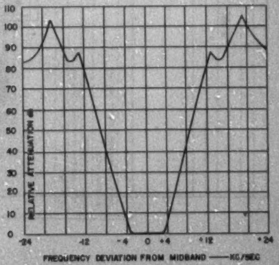


Fig. 6 (right). Receiving band-pass filter. Frequency characteristic

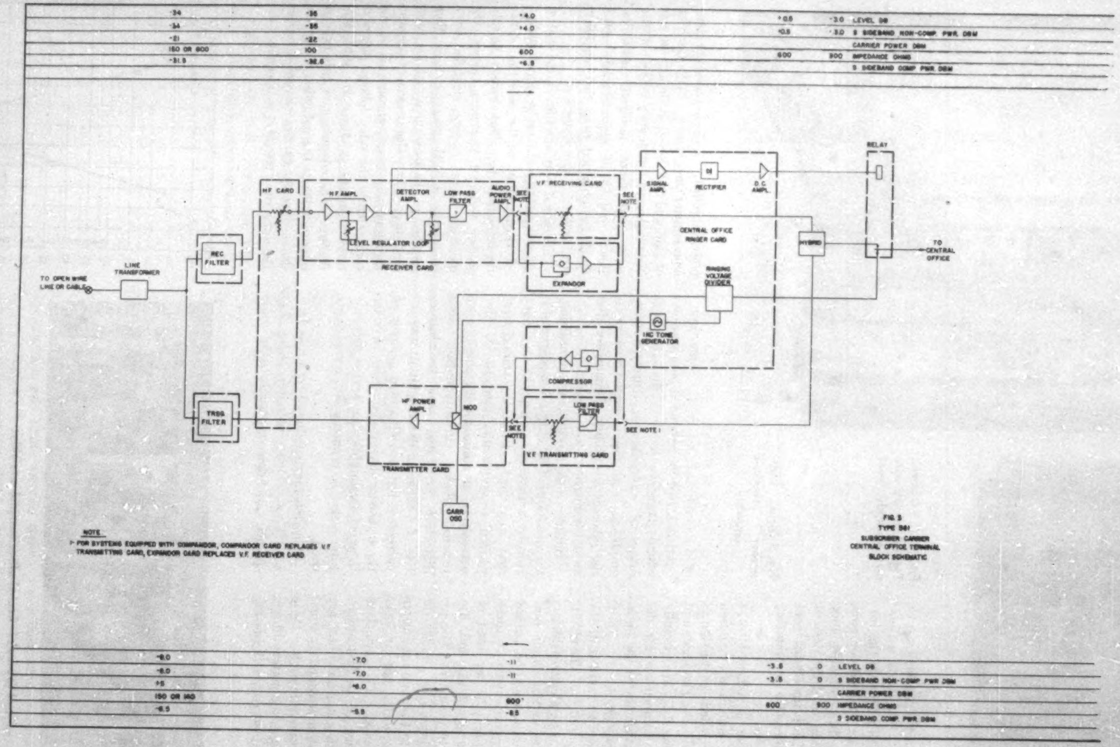

-34
-34
-21
150 OR 600
-31.5
-36
-35
-22
100
-32.6
+4.0
+4.0
600
+6.5
+0.5
+0.5
600
-3.0
-3.0
900
LEVEL DB
S SIDEBAND NON-COMP. PWR. DBM
CARRIER POWER DBM
IMPEDANCE OHMS
S SIDEBAND COMP. PWR. DBM
TO OPEN WIRE LINE OR CABLE
LINE TRANSFORMER
REC FILTER
TRSG FILTER
HF CARD
HF AMPL.
DETECTOR AMPL.
LOW PASS FILTER
AUDIO POWER AMPL.
LEVEL REGULATOR LOOP
RECEIVER CARD
SEE NOTE 1
V.F. RECEIVING CARD
EXPANDOR
SIGNAL AMPL.
RECTIFIER
D.C. AMPL.
CENTRAL OFFICE RINGER CARD
RINGING VOLTAGE DIVIDER
1KC TONE GENERATOR
RELAY
HYBRID
TO CENTRAL OFFICE
COMPRESSOR
HF POWER AMPL.
MOD
TRANSMITTER CARD
LOW PASS FILTER
V.F. TRANSMITTING CARD
CARR OSC
NOTE
1- FOR SYSTEMS EQUIPPED WITH COMPANDOR, COMPANDOR CARD REPLACES V.F. TRANSMITTING CARD, EXPANDOR CARD REPLACES V.F. RECEIVER CARD
FIG. 3
TYPE 561
SUBSCRIBER CARRIER
CENTRAL OFFICE TERMINAL
BLOCK SCHEMATIC
-8.0
-8.0
+5
150 OR 140
-6.5
-7.0
-7.0
+6.0
-5.5
-11
-11
600
-8.5
-3.5
-3.5
600
0
0
900

Fig. 7. Type 561 Subscriber Carrier central office terminal block schematic

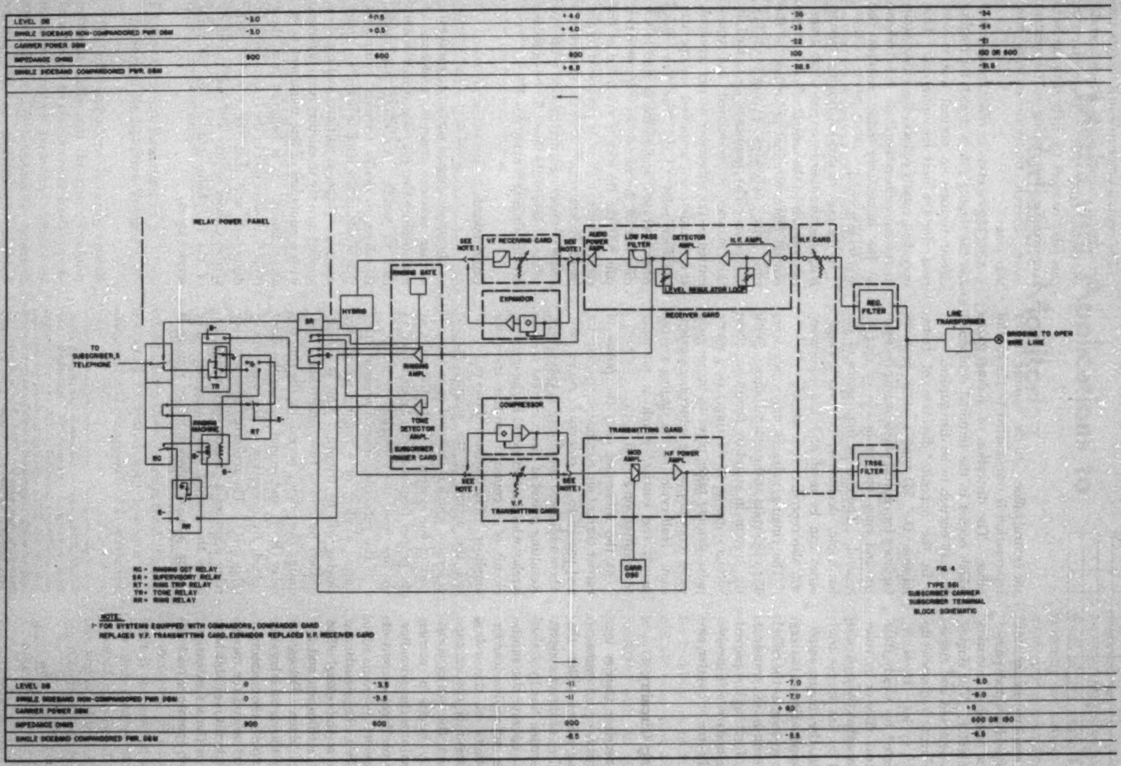

LEVEL DB
SINGLE SIDEBAND NON-COMPANDORED PWR DBM
CARRIER POWER DBM
IMPEDANCE OHMS
SINGLE SIDEBAND COMPANDORED PWR. DBM
-3.0
-3.0
900
+0.5
+0.5
600
+4.0
+4.0
600
+6.5
-36
-36
-22
100
-32.5
-34
-34
-21
100 OR 600
-31.5
RELAY POWER PANEL
TO SUBSCRIBER,S TELEPHONE
TR
RT
RC
RR
SR
RINGING MACHINE
HYBRID
RINGING GATE
RINGING AMPL.
TONE DETECTOR AMPL.
SUBSCRIBER RINGER CARD
SEE NOTE 1
V.F. RECEIVING CARD
EXPANDOR
COMPRESSOR
V.F. TRANSMITTING CARD
AUDIO POWER AMPL.
LOW PASS FILTER
DETECTOR AMPL.
H.F. AMPL.
LEVEL REGULATOR LOOP
RECEIVER CARD
H.F. CARD
TRANSMITTING CARD
MOD AMPL.
H.F. POWER AMPL.
CARR OSC
REC. FILTER
TRSG. FILTER
LINE TRANSFORMER
BRIDGING TO OPEN WIRE LINE
RC = RINGING CUT RELAY
SR = SUPERVISORY RELAY
RT = RING TRIP RELAY
TR = TONE RELAY
RR = RING RELAY
NOTE.
1- FOR SYSTEMS EQUIPPED WITH COMPANDORS, COMPANDOR CARD REPLACES V.F. TRANSMITTING CARD, EXPANDOR REPLACES V.F. RECEIVER CARD
FIG 4
TYPE 561 SUBSCRIBER CARRIER SUBSCRIBER TERMINAL BLOCK SCHEMATIC
LEVEL DB
SINGLE SIDEBAND NON-COMPANDORED PWR DBM
CARRIER POWER DBM
IMPEDANCE OHMS
SINGLE SIDEBAND COMPANDORED PWR. DBM
0
0
900
-3.5
-3.5
600
-11
-11
600
-6.5
-7.0
-7.0
+6.0
-5.5
-8.0
-8.0
+5
600 OR 150
-6.5

Fig. 8. Type 561 Subscriber Carrier subscriber terminal block schematic

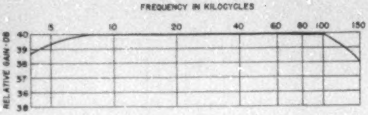


Fig. 9. Transmitting amplifier characteristic

The ringing relay amplifier in the subscriber terminal is driven by signals picked off at the input of the carrier elimination filter following the detector. The ringing relay is operated by detected and amplified carrier pulses at ringing frequencies up to and including 66²/₃ cycles. However, the stage is not affected by steady-state carrier from the central office or by voice modulation. To insure that these circuits will not be actuated by line surges after the ringing has ceased, the ringing amplifier is cut off by the application of battery voltage to the emitter as soon as the called party answers. Because of transient effects, this method of gating has proved superior to other methods which involve the opening of the collector or emitter circuits.

Ringing and Dialing Circuitry

The type *561* Subscriber Carrier terminal provides facilities for transmission of dialing and ringing pulses within the speech channel. Pulses of direct current generated in the subscriber line by action of the telephone dial operate the supervisory relay at the subscriber carrier terminal. The contacts of this relay control the transmission of carrier current. In idle condition the carrier is cut off, and consequently bursts of carrier frequency transmitted to the wire line correspond to pulses of direct current at the subscriber's extension.

At the central office the signals are amplified, and passed to the central office ringer circuit where rectification and d-c amplification takes place. Pulses of direct current operate the signal relay, the contacts of which complete the connection between the central office and the carrier terminal. When the circuit is established, direct current flows in the v-f extension. Pulses of this current operate a relay at the central office. In this manner dialing information is transmitted to the central office. The distortion of dialing pulses which occurs in the terminal equipment is small enough to be disregarded.

In the opposite direction of transmission the *561*-type system provides facilities for central office ringing between tip and ring, and from tip or ring to ground. The ringing frequencies of the harmonic, decimonic, or synchromonic series can be used and the equipment will also reproduce code ringing.

The ringing voltages may vary from 65 to 180 volts at the central office. With the system in stand-by condition, these voltages appear on the ringer card where pulse shaping takes place to insure good pulse repetition at the subscriber's extension. Half-cycles of the ringing voltage switch the carrier, which is normally transmitted, on and off.

At the subscriber end, the pulses of carrier current are rectified in the detector and passed to the subscriber ringer card, where additional amplification takes place. Resulting pulses of direct current operate a relay which is connected in the collector circuit of a common emitter transistor amplifier. The contacts of this relay control the action of the ringing current generator which is included as an integral part of the system. Up to 160 volts at the proper frequency can be delivered to the line from the subscriber terminal.

If the subscriber should answer either during a silent or a ringing period, the ringing will be tripped. This is accomplished by one of two conditions: first, during the ringing period, by the ring trip relay (RT); and second, during the silent period, by operation of the supervisory relay (SR). The action of the supervisory relay operates a ringing gate which disables the ringing relay amplifier. The gating is accomplished before ringing is tripped in the central office thus insuring against line and circuit delays which could cause ringing in the subscriber's ear.

The central office terminal detects the presence of carrier pulses which operate the dialing relay (DR). The contacts of this relay connect the hybrid to the line, thus tripping the ringing in the central office. Speech transmission can now take place.

After the supervisory relay has operated and the ringing gate is closed, the ringing circuit restores to normal and is now ready for the next incoming call. This operation greatly reduces the power drain in the subscriber terminal. The release of the supervisory relay restores the system to normal.

In the case of divided ringing, ground is applied to either tip or ring of the transmission pair at the central office. At the central office terminal grounding of the ring wire activates a low-frequency oscillator. The frequency generated in this manner is modulated and transmitted to the subscriber carrier terminal. It is then detected, amplified, and operates the tone relay (TR). The tone relay, when operated, applies the ground to the ring side of the subscriber's transmission pair. If the tip wire is grounded at the central office, no tone is generated by the oscillator, the relay TR at the subscriber's end is not operated, and ground is applied to the tip side of the line. In this manner, divided ringing is made possible.

Conclusions

The application of carrier to rural subscriber service presents problems of maintenance that may sometimes be difficult to handle. An effort was made, therefore, to design terminal equipment that would insure reliable operation under varying weather conditions and could be applied directly to existing rural facilities, with a minimum of line treatment or rearrangement. It is believed that these objectives were met in the equipment just described, and that the type-*561* Subscriber Carrier System will meet the needs of the telephone industry in the rural service areas.

Radiation-Flux Conditions in Radioactive Media with Applications to Radiation Monitoring

BJORN E. DAHLIN
NONMEMBER AIEE

Synopsis: Calculations are made for the radiation flux in a radioactive medium. The scattering effects are treated with a simplified model giving an easy survey of the total flux in media with various densities and various atomic weights. The calculations are used for comparison of the sensitivities of common detectors for radiation monitoring of air-borne and water-borne radioactivity. The results are in diagrams compared with the stipulated tolerances of radioactivity in those media for the common isotopes.

WE ARE, in many cases, interested in measuring nuclear radiation with its origin in a volume-distributed source. This source may be a nuclear reaction occurring over the volume of a reactor core, radioisotopes present in a reactor-coolant stream, or radioactive gas and dust present in the atmosphere. The purpose of the measurements is generally one of the following three:

1. To determine the hazard of an external exposure to a human being.
2. To determine the health hazard from radioactivity that might expose a human being internally after the radioactivity has been transferred inside the body, normally by inhalation or by eating.
3. To determine a possible failure in nuclear process, as a fuel rupture in the case of nuclear power production with a heterogeneous reactor.

This paper is intended to treat the flux conditions in a few simple cases which may serve for guidance for more complicated practical applications, and also to compare some types of instruments for typical measurements.

Flux of Gamma Radiation in a Spherical Radioactive Volume

The spherical shape is easily treated mathematically and the result can also be modified to give more or less approximate information regarding volumes of other shapes. We assume a spherical volume with an activity of d-emitted radiation particles per second per cm^3 (cubic centimeter). A small spherical volume with a radius r_o in the center of the medium total absorption coefficient μ will be exposed to the number of particles per second ds from each concentric spherical shell with a thickness dr, according to:

$$ds = d(\pi r_o^2 e^{-\mu r})dr$$

Note that ds is independent of r in case of low absorption and therefore all elements give the same contribution supposing their distance to the center is much larger than r_o. If we integrate for all shells over the volume of the sphere with radius R we obtain:

$$S = \frac{d(\pi r_o^2)}{\mu}(e^{-\mu r_o} - e^{-\mu R}) = \frac{d(\pi r_o^2)}{\mu}(1-e^{-\mu R})$$

where the last branch is for $\mu r_o << 1$.

$S/\pi r_o^2$ is the flux F if r_o is infinitely small and therefore

$$F = \frac{d}{\mu}(1-e^{-\mu R}) \qquad (1)$$

The expressions for S and F have to be corrected for two reasons. One is that we have only considered radiation particles that have started with the proper direction for hitting the center volume and have not undergone any absorption process defined by the total absorption coefficient μ. Another contribution not considered, and in many cases very important, is from radiation particles which may or may not have started with a proper direction, but which after one or more scattering collisions will hit the reference volume, radius r_o. We correct for this by introducing a factor B, build-up factor, in the expressions.

Another correction factor, K, has to be introduced in the expression for S if μr_o is not small compared with 1, which was assumed for the expression of ds where the solid angle exposed by the reference element to any point at a distance r was supposed to be $\pi r_o^2/4\pi r^2$. We see that for $r = r_o$ this gives a solid angle fraction of total of 1/4 instead of 1/2, which is evidently the correct value. The maximum error is thus 2 and we can say that K is a factor between 1 and 2.

Our formulas in corrected shape then become

$$S = KB\frac{d(\pi r_o^2)}{\mu}(e^{-\mu r_o} - e^{-\mu R}) \qquad (2)$$

$$F = B\frac{d}{\mu}(1-e^{-\mu R}) \qquad (3)$$

Discussion of the Factor, B

For gamma radiation the most important reason for the build-up effect is the Compton scattering at which point the incident photon loses a fraction of its energy and then continues with a direction determined by the original energy and the energy loss. There are also other reasons for build-up, as bremsstrahlung from recoil electrons, and annihilation radiation from electron-positron pairs. For gamma radiation with an energy below a few mev (megelectron-volts), the case for radiation from alomst all the isotopes with which we are concerned, these other effects are, however, much smaller than the Compton scattering, which is the only case we consider.

The cross-section distribution for various angles of scattering for Compton photons is given by the Klein-Nishina distribution. (The resulting scattering is shown in Fig. 1.)

For a general survey of the build-up effect in our medium, we take the approximate approach of assuming an unidirectional forward scattering of the photons. In the polar diagram it can be seen that higher energies of the photons give a stronger forward dominance. If all absorption is of the Compton type, which is approximately true only in a certain energy range for every element, it is then obvious that the medium behaves as if there were no absorption at all as far as the flux (number of photons per area time) is concerned. For each photon absorbed there is one emitted with reduced energy. The flux then is obtained as the limit value of equation 1 when μ approaches zero and equal to $d(R)$. The corresponding B-factor* then becomes

$$B = \frac{\mu R}{1-e^{-\mu R}} \qquad (4)$$

For higher values of μR the flux and the B-

Paper 57-401, recommended by the AIEE Nucleonic Radiation Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE 2nd Nuclear Engineering and Science Conference, Philadelphia, Pa., March 11-14, 1957. Manuscript submitted March 19, 1957; made available for printing June 19, 1957.

BJORN E. DAHLIN is with the Brown Instruments Division of the Minneapolis-Honeywell Regulator Company, Philadelphia, Pa.

* This factor gives the increase of flux in a volume source which is different from the build-up factor sometimes used for point sources.

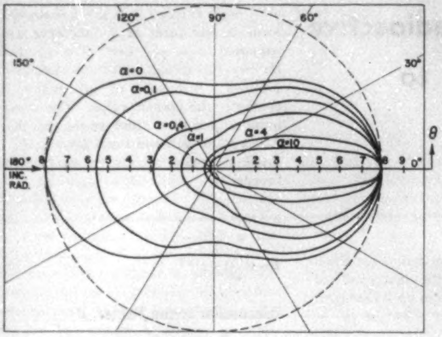


Fig. 1. Differential cross section per unit solid angle for the number of photons scattered at the angle θ (from reference 5)

α = photon energy per electron rest energy
Units—10^{-26} cm^2 per electron

factor must be limited because the photons finally will have, after repeated scattering, an energy so low that they are unlikely to avoid photoabsorption (or possibly be of no interest for a particular measurement because they cannot affect a detector). Equation 4 is thus only valid for a small or a moderate value of μR. B is then comparatively small and the importance of the rude scattering assumption decreases. The energy cut-off, depending upon photoabsorption, can for light elements be regarded as about 60 kev (kiloelectron-volts), but for heavy elements much higher energies must be chosen, (for mercury, for instance, about 0.5 mev).

The average loss of energy of a photon undergoing Compton collisions in water can be calculated from reference 1, and the result used for the derivation of the energies (E_ν) an average photon obtains under successive collisions. The results are shown in Table I where arbitrarily the reference energy has been selected up to 60 kev.

Table I. Energies of Average Photon Under Successive Collisions

Collision No., (ν)	E, (Kev)
0	60
1	65
2	72
3	82
4	99
5	112
6	137
7	173
8	223
9	298
10	420
11	690
12	1,300
13	3,100

The average distance between two collisions of a photon in a medium is $1/\mu_\nu$, where μ_ν is the absorption coefficient at an energy E_ν. When a photon is slowed down in repeated collisions, it will, if we still assume a straight forward scattering, travel an average distance

$$R_o = \sum_{\nu=\nu_c}^{\nu_o} 1/\mu_\nu$$

where ν is the collision index, ν_c is the collision index for the cut-off energy, and ν_o is the index for the original energy. Contributions to the flux at our reference point will only be approximately within a radius R_o. Assuming R_o is much smaller than R (which practically means that $\mu R >> 1$). The flux then is equal to $d(R_o)$ and

$$B = \mu R_o = \mu \sum_{\nu=\nu_c}^{\nu_o} \frac{1}{\mu_\nu} \tag{5}$$

If the absorption coefficient is constant over the energy range considered, B becomes $1+\nu_o-\nu_c$ which is the number of collisions plus 1 (μ with index ν_o is equal to μ).

It is very easy to see that in the case for $\mu R \gg 1$, the type of scattering, unidirectional or not, is of very little influence upon the flux and the B-factor. Let us compare it with another extreme distribution, which is spherically symmetric. The scattered radiation will then act as further volume sources $d_1 \ldots d_n$ where the index is the number of collisions previously undergone. In the neighborhood of the center point, all these virtual sources must be equal to d as the flux from each one is approximately constant and equal to d_ν/μ_ν, where d_ν originates from the source $d_{\nu-1}$, and μ_ν is the absorption coefficient corresponding to the energy. The next virtual source $d_{\nu+1}$ becomes μ_ν times that flux which is equal to d_ν. An integration over the sphere of the nonscattered fluxes from all the sources gives the same result as the foregoing ($B = \mu R_o$).

As the Compton scattering is essentially an interaction between photons and "free" electrons, the average energies for heavier elements should essentially follow the same schedule as given for water. The cut-off energy is higher, however, and as the collision energies are closer at low energies the value of the B-factor becomes rapidly smaller. The simple slowing-down model apparently may give also a rough idea about the energy spectrum that can be expected.

Considerations for Flux of Nonspherical Media

It is convenient to calculate the flux at the surface of the sphere. An integration shows that if no build-up effects are considered, the surface flux will be equal to half the center flux for μR either very large or very small. For other values of μR the flux is also approximately the same. (The value for instance for $\mu R = 1$ is $0.46F$ center.) As the flux has maximum value in the center an approximation of the distribution over the volume may be chosen as a cosine function which has half of its maximum value at a distance R from the center. Thus

$$F = F_{\text{center}} \cos\left(\frac{\pi}{3}\frac{r}{R}\right)$$

The flux from a hemispherical cloud is evidently half the value of that for a sphere if build-up is not considered. If the radius of the hemisphere is infinite, we see that the flux at a plane boundary of an infinite medium is half the value of that deep inside the medium. That will also be true for a finite plane which is large compared with the free mean path, $1/\mu$, and we can assume the foregoing distribution along each center axis of a parallelepiped. (R is equal to half the respective side length.) A spherical quadrant gives, accordingly, at the center point 1/4 of the value of a sphere and 1/8 an octant. The flux in the center of the surfaces of a parallelepiped will then be approximately half the value of the internal maximum flux, double the value of the edge flux, and four times the value of the flux in the corner, still neglecting build-up effects.

When the scattered radiation is considered, the situation becomes less obvious. The scattered flux will depend both upon the angular distribution of the scattered radiation, which is determined by the energy, and upon the surrounding medium. Conclusions can be made for many particular cases, and the upper and lower limits of the scattered radiation can be found by considering the most "optimistic" and most "pessimistic" conditions. If the surrounding medium has about the same atomic number as the radioactive medium, the build-up at the surface should be about the same as in the volume, regardless of energy and densities of the two media. A strong multidirectional scattering distribution in a radioactive medium surrounded by a medium essentially with photoabsorption for the energies present during the slowing down, the build-up at the surface will naturally be much lower than deep inside. For the center point of a hemisphere a spherical symmetrical scattering picture gives the lowest possible value of the B-factor of $(1+B_0)/2$, where B_0 is the value deep inside and can be estimated with the foregoing theory. The value is easily understood as the result of the missing scattering hemisphere. For the center points of quadrants and octants and similarly for edges and corners of a parallelepiped, a similar type of discussion easily can be formed.

EXAMPLE: FLUX IN CONTAMINATED WATER

To illustrate the discussion we should study the flux in a large vessel with water containing a gamma-emitting isotope with an energy 0.5 mev (source strength = d-emitted photons per cm^3 per second). The detector is regarded as capable of responding only to energies above 0.1 mev, and to have a constant detection efficiency over that value up to 0.5 mev. Table I does not exactly fit because our emitted energy is not close to any value in the Table. The average energies under the slowing-down process can easily be derived by a graphical construction of the energy after a collision as a function of the energy before the collision using the values in the Table. The slowing-down sequence can then easily be found graphically by choosing a starting point at 0.5 mev. We do not introduce any important error, however, if we use the energy values in Table I as they are and if we proceed as follows.

The absorption coefficient is at 0.5 mev 0.096 cm^{-1}, and at the tabulated energy values from 0.420 mev to 0.099 mev, 0.10, 0.13, 0.14, 0.15, 0.16, 0.17, and 0.17 cm^{-1} respectively. The effective radius R_e in equation 5 is obtained by adding the reciprocals of the sequence of absorption coefficients. The adjustment for the starting point at 0.5 mev is obtained by adding the reciprocal of 0.096 multiplied by $(0.50-0.42)/(0.69-0.42)$ which is easily understood as an approximation depending upon the position of the incident energy in the interval of the correct sequence. The value for R_e then becomes 52.9 cm. The flux is equal to $52.9d$. The nonscattered flux is $d/0.096=10.4d$. The build-up factor B which is total flux divided by nonscattered flux is 5.1.

If the detector is used at the surface of the vessel the nonscattered flux becomes half the value and equal to $5.2(d)$. The build-up factor B must be lower than 5.1. It must, however, be higher than the value for a hypothetical spherical compton scattering which is $(1+5.1)/2=3$. We can thus say that B is between 3 and 5.1. In this case a sound estimation depending upon the real compton distributions at the average slowing-down energies might be 4.

General Aspects of Radiation Monitoring

In some monitoring cases the volume may have such a low activity that it is difficult to detect, and still represent a hazard of which it is highly desirable to keep a record. The sensitivity of the radiation-detection systems then becomes critical. The sensitivity is for some reasons a vaguely defined property, and for a comparison between different methods we shall have to discuss it briefly.

Many times we want the system to measure one particular type of radiation with a maximum sensitivity and not to be affected by other types. That means that the sensitivity is dependent upon both the detection efficiency and the selection efficiency for the radiation of interest. The former varies for different types of detectors for the same type of radiation, and also varies for different energies of radiation for a particular detector. Several values are given in references 4 and 5. The selection efficiency is far less surveyable as both the system behind the detector, e.g., the electronic parts, and the detector performance have a deciding influence. It is an obvious fact, however, that we cannot detect a particular type of radiation, or call it a signal, which is smaller than the accuracy, A, of the other sources that can affect the system, i.e., other types of radiation, background, or electronic noise. In other words, the minimum detectable signal is the uncertainty of the noise plus background.

Both these phenomenons are statistical in their nature, and if the sources are constant the accuracy with which they can be measured may be taken as the standard deviation. The latter is roughly proportional to the square root of the rate with which the phenomenon takes place. That means approximately that a higher background or a higher noise necessarily gives the system a lower sensitivity. In a good electronic system the noise should represent a constant source and therefore only affect the sensitivity by its statistical properties. For the background radiation we have the same thing, but sometimes we have to consider further large variations also of the average background radiation. Thus the amount of beta and gamma radiation from natural elements in the open field may vary from hour to hour by 25%, and during rainfall there may be an increase of 100%.[3] In monitoring the presence of one particular man-made isotope, naturally, large variations of the background from various activity of other man-made isotopes may be expected.

In the event of a stable background or a background-compensated system the minimum signal, S, at a given background B necessary for an accuracy (relative error), A, (which is between zero and 1) for a time of measurements T, is with an assumption of a Gaussian distribution the following:

$$S=\frac{1}{2A^2T}\left(1+\sqrt{1+8A^2TB}\right)$$

The signal and the background are the rate of incident-radiation particles detected by the detector. The time of measurement, T, is the same for both the signal and the background. The ratio of S/B is apparently decreasing with increasing B or S. It is evident that a comparison between the sensitivity of various systems always must be made with the same integrating time and accuracy to be significant.

The background is usually more important as a limit of the sensitivity of a system than the electronic noise. (Exceptions might be low-energy radiation for a scintillation counter and integrating ionization chambers connected to an electrometer with an extremely high input resistance.) The background normally consists of radiation from possible man-made isotopes and natural radioactive elements present in the earth crust, atmosphere, building material, detector construction material, and cosmic radiation. If no high amounts of background radiation from man-made sources are present, usually the natural radioactive elements in the earth's crust and the atmosphere are most important. Material contamination (particularly alpha radiation) may give, however, for an integrating ionization chamber, a variation of the background reading of the same magnitude as the statistical accuracy of the radiation from the earth's crust. The

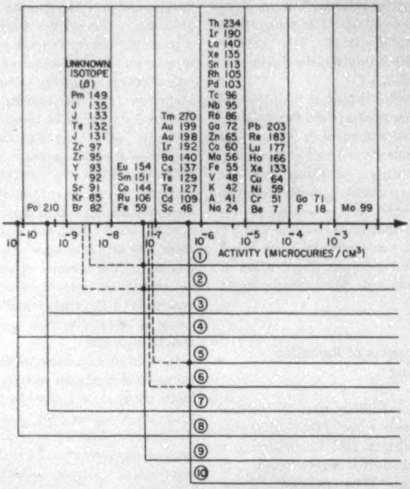


Fig. 2. Occupational tolerance activities in air for some gamma-emitting isotopes (arranged in decade intervals) and comparative sensitivity of some radiation monitors

1. Integrating ionization chamber. Open field
2. Scintillation counter. Open field
3. Filter-tape collector with scintillation counter. Open field
4. Electrostatic precipitator with scintillation counter. Open field
5. Integrating ionization chamber. Building
6. Scintillation counter. Building
7. Filter-tape system. Building
8. Electrostatic precipitator system. Building
9. Equivalent activity of background radiation in open field
10. Equivalent activity of background radiation in a building

gamma radiation from the natural radioelements in the atmosphere is usually much less than that from the earth's crust and building materials. The natural radioactivity in the air, from alpha and beta emitters is, however, an obstacle in air-sample collection systems for monitoring of these two types of radiation.

In the following the sensitivities of common methods for monitoring air and water are compared. There are apparently so many external circumstances and instrumental parameters which are variable that any attempt to treat them in a general way might endanger the surveyability of the subject. An attempt has been made instead to assume what has been regarded as typical. A step away from that might be that the assumed response times or integrating times for some instruments, e.g., ionization chambers, are much longer than often used in practice. The reason for the choice has been to obtain a comparison on equal basis with those systems which are by necessity slow. The term "dip detector" is used to indicate methods where the radiation sensitive element is simply immersed in the medium without any type of sample preparation.

Monitoring of Air-Borne Gamma-Emitting Isotopes in Open Field (Assumed Energy 0.5 Mev)

Integrating Ionization Chamber Used as Dip Detector

We use our derived formulas regarding the source as a hemisphere with a 100-m (meter) radius. The detector is assumed to have a size equivalent to a sphere with the radius 7.5 cm (r_o). The value of μ is 1.11 (10^{-4} cm^{-1}) according to tables, so μR is 1.11. If we assume the ground has the same scattering qualities as the air, the build-up factor will be the same as for a sphere. The formula gives $B=1.65$. The value of μr_o is small and therefore $K=1$.

If we assume $10^{-16}A$ as the minimum measureable current with an accuracy of the order of 5–10% and with an averaging time of about 15 minutes, that will correspond for a normal air-equivalent chamber with one atmosphere pressure to an incident radiation of about 300 photons/second, or a flux of 1.7 photons/cm²/second.

The formula for incident radiation gives when solved for the activity, which in this case is the minimum detectable, $d=1.15\cdot(10^{-4})$ emitted photons/cm³/second. If the isotope has one emitted photon per disintegration that gives a minimum detectable source strength$=0.3$ (10^{-8} microcuries (μc)/cm³).

This detector cannot in this application be compensated for variation of the natural background (supposed average over a time of 15 minutes or more) which is, for operations close to the ground, about $0.005-0.050$ milliroentgen/hour or in flux about $3-30$ photons/cm²/second. The detector then is exposed to about $530-5{,}300$ photons/second. The equivalent volume activity of that source would be about 0.5 (10^{-7} μc/cm³). This is fairly high compared with the tolerance values of several isotopes according to Fig. 2.

Scintillation Counter Used as Dip Detector

We assume a spherical phosphor with a radius of 1.5 cm. We also assume the material is sodium-iodine and can then expect a detection efficiency of about 60%. A natural background radiation of 30 photons/cm²/second which cannot at this energy be electronically separated in a high degree will then give 140 counts/second. With that background we need for a 10% accuracy and an integrating time of 15 minutes, according to the formula given, at least a signal of about 6 counts/second, which corresponds to an incident radiation of 10 photons/second. The minimum necessary flux then is about 1.4 photons/cm²/second. This corresponds for one emitted-gamma ray per decay to 0.25(10^{-8} μc/cm³). It is evident that this device is also limited to variation in the average background radiation which is, as before, about 0.5(10^{-7} μc/cm³). We note, however, that this device is about as sensitive as the integrating ionization chamber.

Continuous Filter-Tape Aerosol Collector with Scintillation Detector

This type of intrument is only sensitive to radioactivity present in the atmosphere as aerosols but not to radioactive gases. The radioactive particles are collected

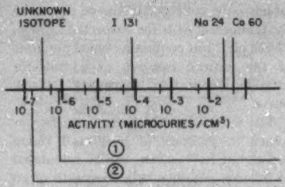


Fig. 3. Occupational tolerance activities in water for some gamma-emitting isotopes and comparative sensitivity of some monitoring systems

1. Integrating ionization chamber with electrometer
2. Scintillation counter

(Fig. 4, right).

1. Geiger counter with separate gamma compensation. (Calculation for average beta energy = 0.5 mev)
2. Filter-tape collector with geiger counter
3. Filter-tape collector with fresh-air compensation
4. Natural background of radon and thoron daughters for the foregoing filter-tape collectors

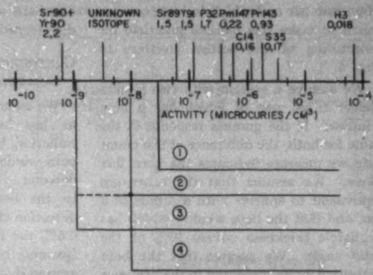


Fig. 4. Occupational tolerance activities in air for some isotopes with pure beta emission and sensitivity of some monitoring systems (maximum energy of radiation in mev indicated below the isotopes)

on filter tape by soaking, with a pump. The filter efficiency is generally about 95–100% for the normal particle sizes of 0.1 by 10 diameter. Pumping speed is generally about 20 cubic feet/hour to 20 cubic feet/minute. The geometrical efficiency of the radiation detector might be 30% of the full solid angle.

We assume the same phosphor as condition 2 in Fig. 2, and the same background radiation. The signal for 15-minute integration time and 10% accuracy then has to be 6 counts per second. If we assume one emitted-gamma photon/disintegration, the sensitivity becomes $3.4(10^{-10}\ \mu c/cm^3)$, (pumping speed 20 cubic feet/minute).

Continuous Electrostatic Precipitator Aerosol Collector with Scintillation Detector

The collection efficiency for this system varies highly with the pumping speed. Maximum collection (i.e., maximum of the product of speed and efficiency) is for a typical instrument at 180 cubic feet/hour where the efficiency is 40%. Assuming the same detector as condition 2 and 3 in Fig. 2, and the same integrating time, accuracy, background, and one photon/disintegration, we get the sensitivity of $0.9(10^{-10}\ \mu c/cm^3)$.

Monitoring of Air-Borne Gamma-Emitting Isotopes Inside a Building

Integrating Ionization Chamber Used as Dip Detector

We assume that we have the same ionization chamber as before, and that the conditions can be represented by a hemispherical "cloud" with a radius of 5 m, surrounded by walls with the same scattering qualities as air. The build-up factor then becomes 1.03. The same conditions for the minimum flux as before of 1.7 photons/cm²/second then corresponds to an activity (one photon/disintegration) of $0.86(10^{-7}\ \mu c/cm^3)$. The equivalent background activity will in this case be larger than in open field and we can assume a value of about $0.5(10^{-6}\ \mu c/cm^3)$, which becomes the limit of the sensitivity.

Scintillation Counter Used as Dip Detector

With the same detector and background conditions as in the open-field measurement we obtain sensitivity of $0.71(10^{-7}\ \mu c/cm^3)$. The equivalent background activity of about $0.5(10^{-6}\ \mu c/cm^3)$ will also in this case be the practical limit.

The sensitivity of filter tape and electrostatic precipitator collector systems will be the same as for the open-field measurements.

Monitoring of Water-Borne Gamma-Emitting Isotopes

Integration Ionization Chamber as Dip Detector

We assume that the medium is represented by a spherical volume with a radius of 0.5 m. The absorption coefficient at 0.5 mev (see Fig. 3) is 0.096 cm⁻¹ and thus $\mu R = 48$. We can use the value of the flux derived in a previous example. There was obtained a total flux of 52.9*d*. For a sensitivity limit of the chamber of 10^{-16} amperes corresponding to an incident flux of 1.7 photons/cm²/second, we can then detect 0.032 emitted photon/cm³/second which is equal to 0.9 by $10^{-6}\ \mu c/cm^3$ for one photon per radioactive disintegration.

Scintillation Counter Used as Dip Detector

We assume the same detector (equivalent sphere with 1.5-cm radius) with the same phosphor as for air monitoring.

The background conditions will be improved, depending upon the shielding effect of the water. We assume the reduction is with a factor 20 to 7 counts per second. With an integrating time of 15 minutes and an accuracy of 10% we then need a signal of 1.3 counts per second. Corresponding flux is 0.31 photon/cm²/second. The minimum detectable activity then is $1.6(10^{-7}\ \mu c/cm^3)$ for one photon per disintegration.

Monitoring of Beta-Emitting Isotopes in Air

Geiger-Counter with Separate Gamma Compensation Tube

An approximate value for the beta flux in the center of a spherical volume can easily be derived in the same way as for gamma radiation. The volume for contribution will be defined by the range of the average energy of the emitted-energy spectrum from the isotope. That range which is generally expressed in grams/cm² has to be corrected because of the wall absorption in the detector by simple subtraction of its area density. Inside this corrected range the medium can be regarded as nonscattering for our approximate calculation purposes. Although this is physically far from true (the beta particles are likely to undergo thousands of collisions with connected probable changes of direction), the symmetry of the medium does make a contribution to a decrease in the error of the assumption.

The wall absorption of a Geiger for beta radiation is usually a few thousands of a gram per cm^2 which is, however, too much to make this method practical for monitoring of the softest emitters in Fig. 4.

We assume a system with two similar Geiger tubes, one of which has a beta window. If the gamma response is the same for both, the difference of the count rate we measure indicates the beta flux alone. We assume that the tubes are equivalent to spheres with a radius of 3 cm, and that the beta window which has negligible thickness covers 10% of the solid angle. We assume that the beta activity is to be measured in the presence of a gamma activity with the same strength (emitted particles per cm^3 per second). The average energy of the beta radiation and the energy of the gamma radiation are both supposed to be 0.5 mev. The radius of the active medium is assumed as 5m. The ratio of the beta flux to the gamma flux then becomes 0.15. If the absorption efficiency of the gamma radiation is 0.5%, for the beta radiation 100%, the ratio of counted beta particles to counted photons becomes 3.0. With an integrating time of 15 minutes and an accuracy of 10%, the signal-background formula gives a necessary signal of at least 0.19 beta particle/-second. This corresponds to a flux of 0.017 beta particle/cm^2/second and an activity of 0.3 (10^{-7} μc/cm^3), assuming one beta and one photon emission per disintegration.

Continuous Filter-Tape Aerosol Collector with Geiger Counter

We assume a similar filter-tape system as has been exemplified for gamma radiation, but using a Geiger tube with a beta window instead of the scintillation detector. If the gamma activity is equal to the beta activity and the gamma detection efficiency has a normal value of 0.5% the ratio of signal to background (gamma) becomes 200 to 1. This background is smaller and negligible compared with the background of beta radiation from collected daughter products of radon and thoron which are present in the normal atmosphere. It is desirable to separate this type of radiation from that of man-made isotopes as many of the latter represent a much higher hazard when present in equal amounts. A normal radon-thoron activity is $10^{-9}-10^{-8}$ μc/cm^3 which thus will be the limit of the sensitivity for the system.

Continuous Filter-Tape Aerosol Collector with Fresh-Air Compensation

There are some methods that can be used to avoid the limitation of sensitivity due to radon and thoron products. One is to operate a detector on a filter tape which only collects air known not to be contaminated with the suspected isotope, "fresh air," but containing equal amounts of the natural isotopes as a collector sampling from the "suspected" atmosphere.

The sensitivity which thus can be obtained in presence of a natural radon activity of 10^{-8} μc/cm^3 can be calculated as follows. With a collection rate of 20 cubic feet/hour we collect during 15 minutes a sample emitting 50 beta particles/second. With a geometrical efficiency of 30% of the Geiger tube the background becomes 15 cycles per second. Necessary signal from the suspected sample is, then, for 10% accuracy and an integrating time of 15 minutes, 1.8 counts per second. The corresponding beta activity is 1.2 by 10^{-9} μc per cm^3, for one beta emission per disintegration.

References

1. Electron Energy Distributions Produced by Gamma Rays, H. E. Johns, J. E. Till, D. V. Cormack. *Nucleonics*, New York, N. Y., vol. 12, no. 10, 1954, p. 40.
2. Area Survey Manual. *BNL 344*, Atomic Energy Commission, Brookhaven National Laboratory, Upton, N. Y., 1955.
3. Radiation Dosimetry, G. J. Hine, G. L. Brownell. Academic Press, Inc., New York, N. Y., 1956.
4. Nuclear Radiation Detectors (book), J. Sharpe. Methuen & Company, Ltd., London, England, 1955.
5. Gamma Ray Absorption Coefficients, C. M. Davidson, R. D. Evans. *Reviews of Modern Physics*, New York, N. Y., vol. 24, no. 2, 1952, p. 79.

Recent Developments in Nuclear Instrumentation at the Knolls Atomic Power Laboratory

R. S. STONE
NONMEMBER AIEE

INSTRUMENTATION in the fields of neutron and reactor physics may be divided into two general categories: The first is the general field of reactor instrumentation. Here, three functions must be fulfilled: the monitoring, the control, and the protection of the nuclear reactor. Until these three functions are fulfilled, the reactor cannot be safely used as either a neutron source or a critical assembly.

After the first or operational objective has been satisfactorily fulfilled, the emphasis of the instrumentation program is shifted to an experimental function. Here, the problems consist of collecting and reducing experimental data to obtain physics information relative to a particular reactor system. Information relative to criticality and control range, power or flux distribution, spectral distribution, and dynamic behavior must all be obtained through the use of nucleonic instrumentation.

A functional diagram of the instrumentation system employed at the critical assemblies and research reactor at the Knolls Atomic Power Laboratory (KAPL) is given in Fig. 1. Since these are strictly experimental reactors, the operational philosophy is to scram the reactor when any one of the monitoring instruments calls for scram even though this signal may arise from equipment malfunction.

The system consists of five ion chambers which are located in proximity to the reactor. These chambers provide a current signal which is proportional to reactor power level for five operational amplifiers.

Three of the operational amplifiers are of a zero-stabilized linear type which provide a multirange linear indication of reactor power level. The ranges are scaled in half-decade steps over a total range of six to nine decades. A nonlinear output stage provides a sharply rising voltage when the linear amplifier output voltage exceeds 1.5 times full-scale indica-

---

Paper 57-399, recommended by the AIEE Nucleonic Radiation and Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE 2nd Nuclear Engineering and Science Conference, Philadelphia, Pa., March 11–14, 1957. Manuscript submitted June 18, 1957; made available for printing July 8, 1957.

R. S. Stone is with the General Electric Company, Schenectady, N. Y.

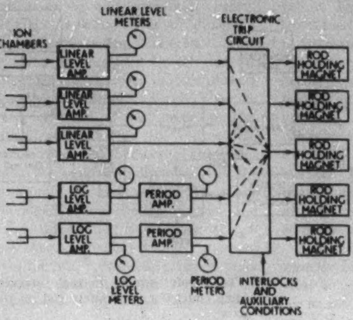

ION CHAMBERS
LINEAR LEVEL METERS
ELECTRONIC TRIP CIRCUIT
LINEAR LEVEL AMP.
LINEAR LEVEL AMP.
LINEAR LEVEL AMP.
LOG LEVEL AMP.
LOG LEVEL AMP.
PERIOD AMP.
PERIOD AMP.
ROD HOLDING MAGNET
ROD HOLDING MAGNET
ROD HOLDING MAGNET
ROD HOLDING MAGNET
ROD HOLDING MAGNET
LOG LEVEL METERS
PERIOD METERS
INTERLOCKS AND AUXILIARY CONDITIONS

Fig. 1 (left). Reactor protective system

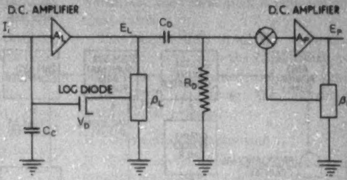

D.C. AMPLIFIER
I_i
A_L
E_L
C_D
LOG DIODE
V_D
C_c
β_L
R_D
D.C. AMPLIFIER
A_P
E_P
β_P

Fig. 3 (right). Logarithmic and period amplifier

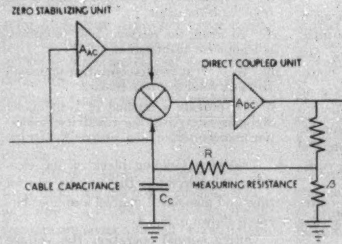

ZERO STABILIZING UNIT
A_{AC}
DIRECT COUPLED UNIT
A_{DC}
E_0
R
MEASURING RESISTANCE
CABLE CAPACITANCE
C_c
β

Fig. 2 (left). Zero-stabilizing amplifier

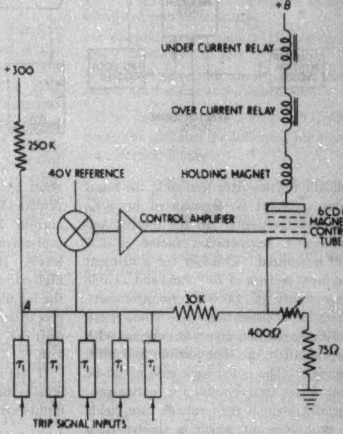

+B
UNDER CURRENT RELAY
OVER CURRENT RELAY
HOLDING MAGNET
6CD6 MAGNET CONTROL TUBE
+300
250K
40V REFERENCE
CONTROL AMPLIFIER
A
30K
400Ω
75Ω
T_1
T_1
T_1
T_1
T_1
TRIP SIGNAL INPUTS

Fig. 4 (right). Magnet control circuit

tion on any range. This signal serves as an input signal to the electronic trip circuit which removes rod-magnet holding current when any one of the five input signals exceeds a preset scram level.

The remaining two operational amplifiers consist of units which provide an output which is proportional to the logarithm of reactor power level and an output which is proportional to the time derivative of the logarithm of power level usually measured in terms of reactor period (*e* folding time). The logarithmic output voltage provides wide range indication of power level on either a meter or recorder, and the period meter provides indication of rate of change of reactor power level. A nonlinear circuit on the output of the period amplifier again provides a sharply rising output voltage if the reactor period becomes unduly short. This increased output voltage removes the holding current for the rod-holding magnets through the electronic trip circuit in the same manner as the signal from the linear amplifiers.

The primary function of the electronic trip circuit is to provide holding current to the rod-holding magnets when all operational conditions have been fulfilled and to remove this holding current from all magnets when any one of the operational conditions is violated. In addition to the five neutron signal conditions previously mentioned, all interlock conditions are fulfilled in this trip circuit.

The over-all system is unique in that the signal path from ion chamber to rod-holding magnet in all cases is a d-c path consisting of amplifiers and passive elements only. This allows one to monitor the d-c levels at specific locations along the signal path to ascertain the present circuit conditions. If these points are within operating limits, then one is assured that the circuit will function properly when a scram signal occurs on one or more of the input amplifiers.

A functional block diagram of the zero stabilized d-c amplifier[1] is shown in Fig. 2. There are two advantages of this type of circuit arrangement over a simpler circuit in which the voltage drop across a measuring resistance is measured with a high input impedance voltmeter circuit. First consider the action of the zero-stabilizing unit. This amplifier lacks fast response; however, its input element consists simply of a vibrating capacitance. The change in the contact potential of this element with temperature can be made extremely small, of the order of microvolts per degree. This change in contact potential appears at the output of the amplifier amplified by a factor of $1/\beta$ which at most, is equal to ten. With normal temperature variations, this amounts to several millivolts change in output voltage which is undetectable on the 10-volt output meter. The contact potential at the input to the direct-coupled unit would appear at the output amplified by a factor of $1/\beta$ without the zero-stabilizing unit. Therefore, if there were a 100-millivolt equivalent drift at the input, this would correspond to 10% full scale deflection of the output meter. With the zero-stabilizing unit in the circuit, this drift is attenuated by a factor of A_{ac} or approximately 2,000. This reduces the change in output voltage from the assumed 1 volt to 0.5 millivolt which again is undetectable on the output meter. The stabilizing unit therefore provides an amplifier that has no detectable zero drift.

The second advantage is that by ap-

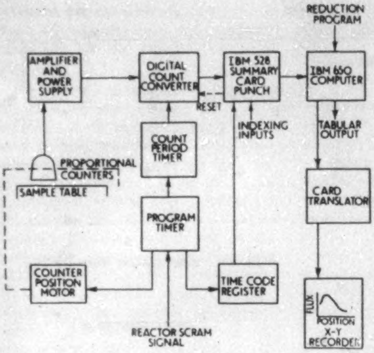

REDUCTION PROGRAM
AMPLIFIER AND POWER SUPPLY
DIGITAL COUNT CONVERTER
IBM 528 SUMMARY CARD PUNCH
IBM 650 COMPUTER
RESET
COUNT PERIOD TIMER
INDEXING INPUTS
TABULAR OUTPUT
PROPORTIONAL COUNTERS
SAMPLE TABLE
PROGRAM TIMER
CARD TRANSLATOR
COUNTER POSITION MOTOR
TIME CODE REGISTER
REACTOR SCRAM SIGNAL
FLUX
POSITION
X-Y RECORDER

Fig. 5 (left). Activation data-reduction equipment

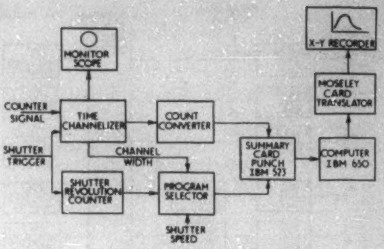

X-Y RECORDER
MONITOR SCOPE
MOSELEY CARD TRANSLATOR
COUNTER SIGNAL
TIME CHANNELIZER
COUNT CONVERTER
SHUTTER TRIGGER
CHANNEL WIDTH
SUMMARY CARD PUNCH IBM 523
COMPUTER IBM 650
SHUTTER REVOLUTION COUNTER
PROGRAM SELECTOR
SHUTTER SPEED

Fig. 6 (right). Spectrum data-reduction equipment

plication of negative feedback, the input time constant is degenerated by βA_{dc} for fast transients. In the present amplifiers, a degeneration factor of 2×10^4 is realized. Consider for a moment an input resistor of 10^{11} ohms and a cable capacitance of 1,000 micromicrofarads. Without degeneration, this would yield a 100-second time constant, whereas with degeneration, this time constant is 5 milliseconds. The unfed back case would be completely unacceptable for reactor protection, whereas the feedback case yields a time constant which is shorter than necessary in most cases. This negative feedback therefore provides one with a reactor control instrument that is sufficiently fast on all ranges to protect the reactor adequately.

The trip unit consists of a negatively fed back loop which maintains a constant output voltage of 40 volts d-c over the full-scale range of the instrument. When the amplifier output exceeds 1.5 times full scale, the negative feedback in this unit is removed and the trip output voltage rises rapidly to 100 volts. This change in output voltage is applied to the electronic trip circuit as a scram signal.

A functional diagram of the logarithmic and period amplifier[2] unit is shown in Fig. 3. Consider an input current from a power-level-monitoring ionization chamber. The output voltage of the log amplifier E_L is given by V_D/β_L under the assumption that $A_L>>1/\beta_L$. At low currents, a diode has a characteristic that $V_D=K \log I_t$ over a range of six to seven decades in current. Therefore

$$E_L=\frac{K}{\beta_L}\log I_t$$

The effect of negative feedback over the diode is to degenerate the input time constant by a factor of $A_L\beta_L$. This again provides fast output response at low input current levels, thereby providing fast protection to the reactor at low power levels. It is possible to connect a second high-current diode in series opposition to the log diode in order to obtain some degree of compensation for contact potential drift due to diode filament voltage variation.

If one assumes that the reactor power level is increasing exponentially, then the input current will be of the form

$$I_t(t)=I_0\epsilon^{t/\tau}$$

The voltage at the output of the log amplifier is then

$$E_L=\frac{K}{\beta_L}\left(2.3\frac{t}{\tau}+\log I_0\right)$$

When the voltage is applied to the period amplifier, its output will be

$$E_P=\frac{2.3KR_DC_C}{\beta_L\beta_P}\frac{1}{\tau}\left(1-\epsilon^{-\frac{t}{R_DC_D}}\right)$$

Asymptotically, this output voltage is then inversely proportional to the reactor period. It is therefore possible to calibrate a meter in units of reactor period.

The trip amplifier stage in the log and period amplifier unit consists of a single pentode stage which is normally drawing grid current. The plate voltage is thereby maintained at a voltage of approximately 40 volts. As the trip period is approached, the grid of this tube is carried negative and the plate potential rises rapidly. This potential rise trips the electronic trip circuit, thereby shutting down the reactor.

The electronic trip circuit[3] in use at KAPL is designed with maximum reactor safety as a basic consideration. The functions of this circuit are:

1. To supply current to each magnet-holding clutch for each safety rod on the reactor.
2. To remove all holding current when any one of the 3-level or 2-period inputs call for reactor scram.
3. To indicate which of the five inputs was responsible for scram.
4. To scram the reactor under conditions of equipment malfunction.
5. To prevent reactor operation unless all interlock conditions are fulfilled.
6. To provide a method of individually dropping safety rods for reactivity calibration experiments.

The basic building block of the electronic trip circuit is the magnet control tube and associated control amplifier circuit is shown in Fig. 4.

During normal operation, the voltage at each of the trip signal inputs is 40 volts; therefore there is no current flow through the Thyrite resistors r_1. The control amplifier acts to hold the voltage at the *6CD6* cathode constant and therefore maintains constant magnet current. As one of the trip input voltages rises above its normal level, current flows through the Thyrite resistor, and the resistance is sharply decreased. This essentially opens the feedback loop and the magnet current is now controlled by the trip input signal. As the trip input voltage increases, the input voltage to the control amplifier rises sharply. This results in an increasing bias on the *6CD6* and a sharp reduction in magnet current, thereby releasing the rods.

All auxiliary conditions which must be fulfilled prior to operation are imposed by the main interlock circuit. They are:

1. At least one log and period] amplifier circuit must be connected into the trip circuit.
2. At least two linear level instruments must be connected into the trip circuit.
3. All manual scram switches must be in the circuit.
4. All entrances to the reactor cell must be closed.

In concluding the discussion on

reactor safety and monitoring instrumentation, a listing of the advantages of this particular system is in order. First, by using vacuum-tube circuits for the primary signal transfer devices, reactor shutdown is accomplished with maximum speed. By use of monitor meters at critical points within the circuit, the present operating condition of the equipment is constantly displayed for the operators. Allowing each input to scram individually each holding magnet, and then letting the first holding magnet feedback through auxiliary circuitry to scram all of the rest, provides a maximum probability of scram and greatly reduces the dependence of the proper functioning of the circuit on individual circuit components.

The second area in the field of nuclear instrumentation to receive attention at the KAPL is that of instrumenting reactor and neutron physics experiments. Here the problem is to accumulate and reduce a large amount of experimental data. In the following discussion, two particular instrumentation systems are described which greatly reduce the amount of routine calculation in data reduction.

A critical assembly is useful as a tool for determining the power generation rate as a function of position throughout the reactor core under different operating conditions. This information is necessary in order to design a reactor for efficient fuel utilization and to design the reactor cooling system. This type of experiment is accomplished by activating fuel strips or foils at various points in the reactor core and plotting relative activation as a function of position. A core with a heterogeneous structure requires a very detailed study which involves a large-scale accumulation and reduction of data. An automatic neutron flux scanner has been developed at the KAPL by Frost and Fehr.[4] A functional block diagram of this equipment is shown in Fig. 5.

Three long, thin samples are irradiated in different positions in the critical assembly. These samples are then mounted on the scanner table. A programmed position servomechanism carries three shielded beta counters down the length of the activated samples, stopping at uniform increments of length for predetermined count accumulation times. At each position, the accumulated count is printed out on an International Business Machines card using a count converter and the *IBM-523* summary card punch. The deck of cards generated is then used as an input deck for the *IBM-650* computer. The computing program subtracts the background counting rate, and makes corrections for activation time and the decay of activity due to the time delay between generation rate as a function of position along the activated samples. By activating a number of samples at various points in the core, it is then possible to to obtain a complete flux map for the reactor. The output data are available in tabulated form or they can be plotted directly by the use of a Moseley card translator and 2-axis recorder.

A second type of experiment that has recently been fully instrumented is the measurement of neutron energy spectra by means of a neutron time of a flight spectrometer.[5] A neutron beam is brought out of a reactor or a subcritical assembly excited by a fission source. This beam is chopped into a series of short bursts of neutrons by a rotating shutter. After a given flight path length, the neutrons are detected in a counter. The outputs counts are sorted as to time of arrival at the counter by a time channelizer. Since the higher-energy neutrons have a larger velocity, they require a shorter time to traverse the distance between the shutter and the counter than a lower-energy neutron. Therefore, the channelizer gives the velocity distribution of neutrons in the beam. These data can be corrected for counter efficiency, system transmission and background and then transformed into an energy spectrum corresponding to the spectrum at the source point within the reactor.

The channelizer is a recent development of Davidson and Wade[6] of KAPL. Its magnetic core memory consists of 256 individual channels each with a count storage capacity of 2^{16} counts. The unique feature of the design is that the circuit is entirely transistorized. The comparable vacuum-tube equipment consists of four racks of equipment, where the transistorized version occupies a single 10-inch rack panel.

The channelizer is easily adapted to an automatic data-processing system, as shown in Fig. 6. The channelizer stops accumulating data when any single channel is filled. A readout system is provided in which the stored count in each channel is converted to a series of pulses. These pulses are fed over a coaxial transmission line to a computing facility. Here they are converted to a decimal representation by a count converter whose output goes to an *IBM-523* summary card punch. In addition to the counting information, auxiliary information as to the number of shutter revolutions, the counting channel width, and the run number is punched on these cards.

The output deck from the summary card punch is fed into the *IBM-650* computer with a prepunched computing program deck containing the following instructional material:

1. Subtraction of background.
2. Correction for the energy dependence of system transmission including counter efficiency and materials in the beam.
3. Calculation of the mean neutron energy associated with each channel.
4. Transformation from a time-of-flight spectrum to a flux/*eV* spectrum.
5. Correction relating neutron beam current to neutron flux at the source point.

The output deck from the *IBM-650* computer contains the following information for each channel:

1. Mean energy associated with that channel.
2. Flux per electron volt at this mean energy.
3. Probable error associated with this flux determination.
4. Logarithm of the mean energy.
5. Logarithm of flux per electron volt.
6. Logarithm of the flux plus and minus the probable counting error.

The cards in the output deck may be used to tabulate the output or to feed a card translator which plots the spectrum on log-log paper with plus-and-minus probable errors.

The saving in man-hours with the use of automatic data reduction methods in cases like the two cited is obvious. The greater advantage in the application of this type of instrumentation comes in freeing the professional person from the routine task of data reduction and presenting his experimental results in a much shorter time than previously required.

References

1. An Improved D.C. Amplifier for Reactor Control, E. J. Wade, R. S. Stone. *Nucleonics*, New York, N. Y., Apr. 28, 1955, pp. 28–30.

2. A Log and Period Amplifier for Reactor Control, R. S. Stone, E. J. Wade. *Report 1236*, Knolls Atomic Power Laboratory, Schenectady, N. Y., Oct. 15, 1954.

3. An Electronic Trip System for Reactor Protection, E. J. Wade, D. S. Davidson. *Report 1528*, *Ibid.*, Jan. 17, 1956, pp. 84–87.

4. Automatic Neutron Flux Scanner, R. T. Frost, R. B. Fehr. *Nucleonics*, Mar. 1956.

5. Reactor Spectrum Measurements Using a Neutron Time of Flight Spectrometer, R. S. Stone, R. E. Slovacek. *Report 1499*, Knolls Atomic Power Laboratory, Mar. 15, 1956.

6. Transistorized Time of Flight Analyzer with Ferrite Core Memory, E. J. Wade, D. S. Davidson. *Convention Record*, Institute of Radio Engineers, New York, N. Y., Mar. 1957.

Electrical Problems in Electrical Burn-Out Testing of Nuclear Fuel Elements

T. W. HUNT
ASSOCIATE MEMBER AIEE

Synopsis: Burn-out testing involves the heating of reactor fuel elements to destruction or near destruction. Its purpose is to provide a knowledge of the upper limits of power that can be removed from a nuclear reactor. The heating is done electrically because the electric heat can be conveniently applied and accurately controlled and measured. Also, the resulting heat distribution is very similar to that due to nuclear action.

Electrical problems are associated with every phase of the testing and most of them have some nonstandard feature. The paper gives a detailed description of the electric equipment used in a burn-out test facility and discusses solutions to the various problems that have arisen during the test program.

IN pressurized water-cooled reactors, the nuclear fuel elements are a highly concentrated power source. It is possible that the amount of heat released can become so high that the water cannot remove this heat as fast as it is generated within the fuel element. This phenomenon is called heat-transfer burn out and will cause the failure of the heated fuel element due to excessive temperatures.

In water-cooled reactors, burn out occurs because of a change in the water heat-transfer properties. As shown in Fig. 1, the heat transfer passes through several phases.[1] The nucleate-boiling and film-boiling phases are those important in burn out. As the nuclear power level increases, the heat removed from the surface of the fuel element increases and a condition is reached where the heat-transfer action in the water changes from nucleate boiling to film boiling. As can be seen, the wall temperature for film boiling at the same power rate is in the failure region for the fuel element.

Extensive experimental programs are being carried out to determine the upper limits of power that can safely be removed in a given fuel element. These programs generally depend upon electric heating of test specimens which are similar or equivalent in structure to the nuclear fuel elements used in pressurized water reactors.

Perhaps the most significant work in this field is electrical burn-out testing, in which test specimens are heated at current densities of tens of thousands of amperes per square inch. In burn-out testing, electric heating is ordinarily used instead of nuclear heating because of the ease and accuracy with which electric power can be applied and measured. Furthermore, the resulting heat distribution in the metallic element resembles that produced by nuclear action.

A typical test illustrates the effectiveness of the electrical method of burn-out testing. A plate assembly, somewhat similar to the one discussed by Cunningham and Boyle,[2] when tested to failure, failed as shown in Figs. 2 and 3. Such a failure might render a testing reactor inoperative and certainly would result in complete failure of an operating power reactor. In a reactor, instrumenting such a test would be difficult; moreover, the time and cost would be prohibitive. On the other hand, the use of the electrical burn-out method makes testing the assembly shown in Fig. 2 safe, easy, accurate, and relatively inexpensive.

General Description

The test facility was designed to produce all of the conditions leading to heat-transfer burn out that a given fuel element will experience in any water-cooled reactor. The facility includes a low-voltage high-current power supply, and several hydraulic test loops and their associated instrumentation and control. Fig. 4 shows a complete testing unit consisting of a 300-kw power supply, hydraulic test loop, and control panel. Fig. 5 shows this same testing unit schematically.

The power supply consists of a separately excited, shunt d-c generator, with fine field control over the entire voltage range from residual to about 45 volts. It can develop up to about 16,000 amp (amperes) for short periods.

The facility has several testing units like the one shown in Fig. 4. Up to the present, only one d-c power supply has been available. A larger d-c power plant is now being built which will have a continuous rating of about 25,000 amp at 115 volts. This power plant will include a 4-unit motor-generator set consisting of one 4,310-horsepower 13-kv synchronous motor driving in tandem three 750-kw generators. Special features will be provided, such as arrangements for operating the generators in parallel on large loads and singly on small loads. Also, the control is arranged for obtaining very fine voltage control over the entire range from residual to maximum.

The hydraulic test loop shown in Fig. 5 is typical and includes canned motor pumps, electric heaters, electrically heated pressurizer, instrumentation and control, and ion exchangers, as well as the usual heat exchangers, valves, and high-pressure stainless-steel pipe capable of containing water at up to 2,500 pounds per square inch absolute and 650 F (degrees Fahrenheit).

As shown in Fig. 4, the control panel includes a number of precision recorders. The important variables measured are power to the test section, test section cooling-liquid flow, temperatures of the fluid and test section at various points, fluid pressure, and fluid purity.

Problems of Test Facility Design

In the test facility described, the components must be chosen with considerable care if many different types of tests are to be performed successfully.

HEATING

Reactor fuel element heating must be duplicated as closely as possible. Since reactor elements generate large amounts of heat, the electrical heat source in the test facility becomes an important item. It must be capable of heating prototypes of any of the various fuel element geometries that may be used in the reactor. For example, one typical test section may require 14,000 amp at 40 volts.

CURRENT

In planning the test facility, various tests on many test sections such as the one described were considered, and the separatey excited d-c shunt generator

---

Paper 57-404, recommended by the AIEE Nucleonics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE 2nd Nuclear Engineering and Science Conference, Philadelphia, Pa., March 11-14, 1957. Manuscript submitted October 4, 1956; made available for printing July 11, 1957.

T. W. HUNT is with Westinghouse Electric Corporation, Pittsburgh, Pa.

The author wishes to express his gratitude to the following: Westinghouse Electric Corporation and the U. S. Atomic Energy Commission for permission to publish this work; P. N. Ross, S. J. Green, and Patricia Brown for their very helpful suggestions. The author also wishes to thank the following co-workers for their co-operation: S. W. Cota, R. A. De Bortoli, H. S. Jacket, J. D. Roarty, and A. Weiss.

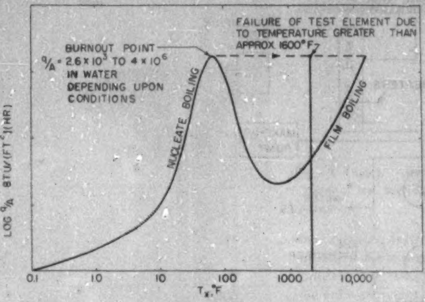


Fig. 1. Boiling heat transfer and burn out

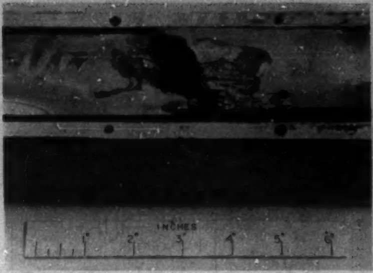


Fig. 3. Burn-out area in multiple-plate assembly

was selected. There are thermal, hydraulic, and electrical reasons for using direct current rather than alternating current. The thermal reasons are not too well defined nor is there complete agreement about them among experimenters. Tests have been run using both types of current, and the results are in fair agreement.[3] However, analysis by the University of Southern California[4] of its heat-transfer work shows that the formation of steam bubbles might be affected by the cyclic nature of alternating current.

Alternating current produces vibration between adjacent test elements in parallel-channel tests because of the cyclic nature of the forces between conductors. Two parallel conductors carrying currents in phase experience attraction; these attractive forces vary between zero and a maximum as the currents vary between zero and a maximum, plus or minus. Measurements have been made which show that the amplitude of the vibration so induced for flat plates unsupported at the edges may be greater than 10% of the flow-channel width. This vibration produces alternate high and low water-velocity pulses, and, since the burn-out flux is a function of the flow velocity, different burn-out fluxes might be obtained through use of alternating rather than direct current. Of course, channels with rigid-edge supports cannot vibrate appreciably.

If the test section thermal capacity is small, alternating current causes alternating wall temperatures which could also cause burn out at unexpectedly low average power.

Alternating current might require excessively large equipment because of the inductive drop at high currents and low test section resistance. The ratio of inductive reactance to resistance can be unfavorable. Where very high currents are involved, very poor power factors result, unless the test section is built into the transformer terminals. This would further complicate test work already extremely difficult.

Power Supply

D-c power can, of course, be supplied by d-c rotating generators, metallic or mercury-vapor rectifiers, or mechanical rectifiers. The factors that influence selection of the power source are smoothness and fineness of voltage control from nearly zero to maximum, ability to develop voltage transients, ability to take overloads, and reliability. Fine, smooth voltage control is essential in obtaining accurate burn-out data. The approach to burn out is usually made very slowly while accurate readings of flow, temperature at many points, pressure drop at many points, current, and volts, etc., are taken. Since peak loads at burn out may be of short duration, machines with inherent large overload capacity can sometimes provide power for burn-out tests above their continuous rating, and can

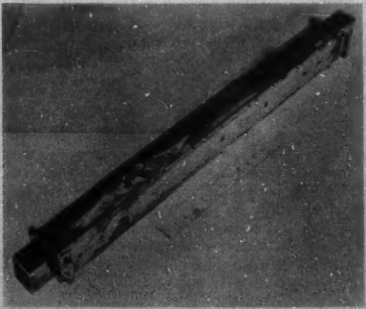


Fig. 2. Multiple-plate assembly after burn-out test

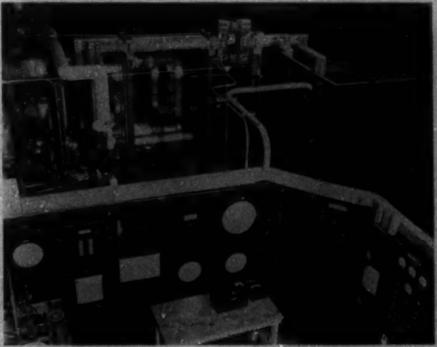


Fig. 4. Burn-out test facility

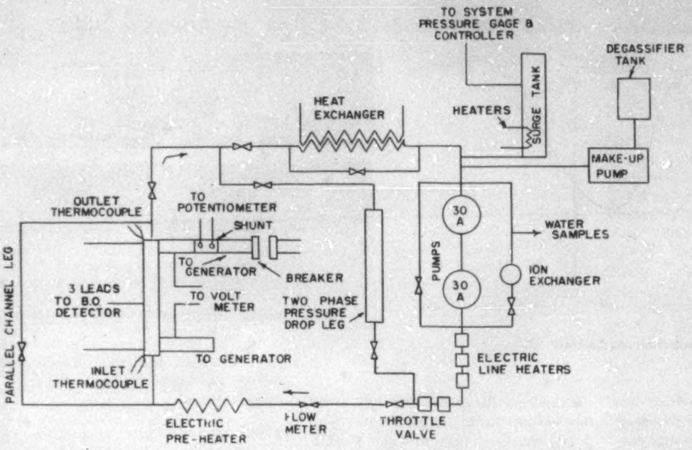


Fig. 5. Schematic diagram of loop

therefore be used as cheaper or more versatile power sources.

Where heat transients are desired, the rapidity of voltage response may be a deciding factor. In the past, attempts have been made to procure equipment that could double power in 0.1 second. Large rectifiers were recommended rather than d-c generators because the fields of d-c generators were not expected to respond fast enough.

The power supply, whether a rectifier or rotating equipment, has to be equipped with a smooth voltage control from essentially zero to the maximum. In the test facility described, the power supply is a generator having field control with separate excitation. A main and vernier rheostat provide essentially stepless concontrol from residual to the maximum voltage obtainable from the machine.

Hydraulic Components

The various hydraulic components must be carefully selected and oriented with respect to each other. Fig. 5 shows that the loop contains two 30-gallon-per-minute canned motor pumps. These pumps normally circulate the hot water at 2,000 psig (pounds per square inch gauge) and up to around 610 F. Location of the heat exchanger between the test section and the pumps prevents them from receiving the full temperatures developed in the test section.

Canned motor pumps are used because of their ability to withstand temperatures pressures, and thermal shock, as well as because of their leakproof qualities. The motor is integral with the pump. Its stator windings are water-cooled and the rotors and bearings operate completely immersed in the high-pressure water. Such pumps have survived 3 years of testing without appreciable maintenance, although pumping water at temperatures near the 636 F boiling point.

Pressure is maintained by using immersion electric heaters to produce steam in a vertical pressure vessel. The tank is uninsulated so as to obtain good transient response in the control of moderate pressure surges.

Electric heaters and a heat exchanger are used to balance the heat generation in the test section and thus keep the loop water temperature constant.

Valving is provided to permit control of the liquid flow, removal of various components, maintenance, etc., as required. Smooth throttling control for a wide range of flow rates is included.

An ion-exchange system is provided to maintain high water purity. A purity corresponding to a resistivity of approximately 1 megohm per cubic centimeter is maintained for several reasons. The heat-transfer surfaces must be kept clean so that results will be uniform. Conduction of electric currents through paths other than the test section must be reduced. Electrical corrosion is a function of the electric leakage current. The higher the water resistivity as well as the greater the length of path, the lower the leakage current. Tests have been made where electrically induced corrosion, even with 1/2-megohm water, pitted and otherwise seriously corroded potential test sections.[5] This problem can be avoided by a design that will prevent the existence of high-potential gradients through the water or limit the gradients to areas of no consequence.

Controls

The loops are instrumented so that accurate test conditions can be maintained. Pressures are controlled to within 1% of the desired value through use of a pressure recorder controller. The liquid level in the surge tank is measured, and an automatically operated make-up pump maintains the desired level. The ion-exchanger flow is monitored. The main variables of fluid and test section temperatures are continuously monitored and recorded. These instruments are of the highly accurate, balanced-potentiometer type and all have a fairly rapid response in order to record low-speed transients. Occasionally, where more rapid response time is required, oscillographs have been used. Fig. 4 shows the instrument and control panel used in burn-out loop no. 1.

Test Sections

Although the loop provides some problems due to the operating pressure, variations in operating temperature or temperature transients, and accuracy of control, the main interest and problems are associated with the test section.

Each test section has its own problem dependent upon geometry. Fig. 6 shows one of the first test sections used in the test facility. The test part of this section is a nickel tube 0.187-inch inside

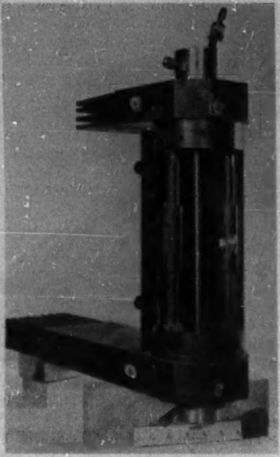


Fig. 6 (left). Round-tube test section assembled in housing

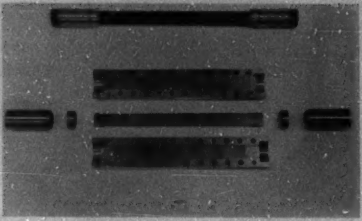


Fig. 8 (right). Rectangular-channel test section design

diameter and 0.020-inch wall with a cross section of 0.013 square inch. Currents as high as 2,250 amp have been passed through this tube. This is a current density of about 173,000 amp per square inch.

TERMINAL DESIGN

Since burn-out measurements are made on the tube, occurrence of burn out in the connecting leads or terminal is undesirable. A large block of copper is used for each terminal. The nickel tube passes through the center and is brazed to the copper. Since the copper terminal thickness is perhaps 50 to 100 times the tube thickness, there is no point where the current density comes very close to the current density in the nickel. Also, the copper terminal is water-cooled similarly to the test portion. Test results indicate that the large copper block cools the tube sufficiently to cause burn out to move down the tube about 1/4 to 1/2 inch.

During the approach to burn out, the nickel tube may reach a wall temperature of 700 to 900 F. The tube expands lengthwise, and because of its small strength as a column it must be free to move. Expansion of the terminals relative to the bus is permitted by flexible shunts or cables attached between the copper terminals and the high-current bus, see Fig. 7. One end of the test section is attached rigidly to the loop; the other is attached through a looped section of tubing. In this particular arrangement turnbuckles and springs are used to place the nickel tubing initially in tension.

A later design proved simpler and more satisfactory. Although the current is conducted into the tube in the same way, the copper terminal is a cylinder, similar to that shown in Fig. 8. The main terminal is clamped to the smaller copper cylinder. The heavy terminals are spring-supported and the turnbuckles eliminated, which facilitates assembly and disassembly of the test section.

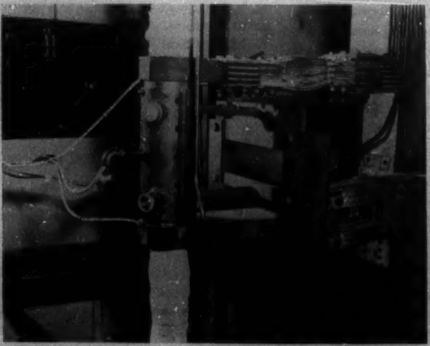
INSULATION

To prevent current from circulating through the loop piping it is necessary to insulate the piping from the terminal at one end of the section. To accomplish this a pair of flanges was machined to receive an insulating gasket, and the bolt holes were machined to receive insulating gaskets around the bolts and under the heads. The gasket material used was compressed asbestos. At one time mica was tried; however, it did not seal well when subjected to pressure and temperature. The compressed asbestos withstands the high temperatures, takes high bearing loads, seals very well, and is easy to cut and form. Consequently, it is a favorite at these low voltages where seals are also required.

CHANNEL DESIGN

Many tests are being performed with channels constructed as shown in Fig. 8. In these cases, the cylinder-type terminal with clamped joint is easiest to use. By making the cylinder sufficiently large, almost any desired ampere capacity can be obtained.

In some cases, tapered entrance channels are required. The cylinders are cut axially, hollowed out as desired, rewelded together, then machined. Fig. 8



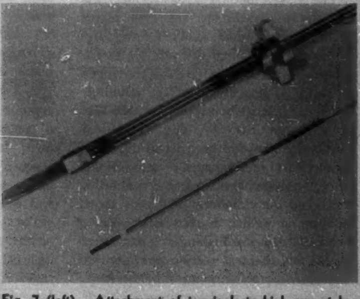


Fig. 7 (left). Attachment of terminals to high-current bus

Fig. 9 (above). Parallel-flow-rod test assembly

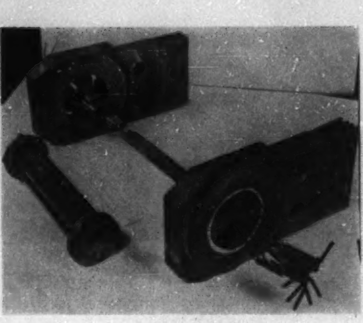


Fig. 10 (left). Terminal and water-seal arrangement of rod cluster

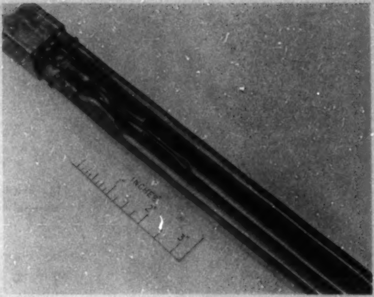


Fig. 11 (right). Burn out in parallel - flow - rod test assembly

shows some of the details of a joint for a zirconium rectangular section. A number of problems have been worked out in the construction of this particular section.

Construction of the channel proved difficult. The channel consists of two parallel plates joined at the edges by a thin member. The heat generated in the plate is greater than that in the thin member because of the higher total current; therefore, burn out will occur in the plate area rather than in the corner.

Since total heat generation in a given area is a function of metal thickness, measurement of burn-out heat based on average current will be in error if the plate wall thickness is not held constant to a close tolerance. About 1-mil maximum variation has proved suitable in most of the test work.

The accuracy of burn-out tests is also affected by the dimensions of the water-flow channel. The burn-out flux is a function of the flow velocity and channel thickness. Consequently, the channel dimension must be maintained accurately. Various methods were tried before a successful one was found.

The present method of making channels starts with machining the side plates separately to prepare them for welding. Following machining, the side plates are assembled on a mandrel and the edges welded. After removal of the mandrel, the channel must be attached to the terminals; see Fig. 8. The cylindrical terminal is made in two parts. First, a short section of the cylinder is machined to receive the channel and the channel is welded to the cylinder. Then the short cylinder is welded to the longer part. Finally, the over-all cylinder is machined to size and the test section is ready for leak testing prior to going into the loop.

Because these test sections will not withstand 2,000 psig without distortion, they must be backed up with heavy back-up housings. Of course, these housings must be insulated from the test section or they would effectively short-circuit it. Compressed asbestos is used as an insulator because of its resistance to high temperature and the ease with which it can be handled.

The clamped joint and cylindrical terminals permit the use of zirconium in the test section rather than nickel or some other substitute metal. Successful methods have not been developed for brazing or welding zirconium to copper or to a large number of other metals. By constructing the complete test section, including the cylindrical terminals, from zirconium, it is possible to eliminate the need for joining copper to zirconium.

Testing a rather close cluster of round tubes presented several new problems. The tubes were cooled on the outside; i.e., the water flowed through the cluster. The terminals had to go into the pressure vessel and had to be sealed to the housing so that 2,000 psig water at 600 F did not leak out around them. At the same time, they had to be insulated from the housing. The test section had to be attached to the terminals inside the housing, allowing provision for differential expansion while providing a seal against the pressure. Also, very little obstruction could be allowed in the water channels. Figs. 9 and 10 show the cluster test assembly developed.

The terminals are constructed of a large, thick copper plate. Asbestos gasketing insulates the terminals from the flanges and the flanges from each other. One end of the test section is permanently attached to a copper cross, bolted to the terminal. The cross provides water-cooled contacts having 6 square inches of surface to carry 12,000 amp. The cross also ensures even distribution of amperes to the test section tubes.

The other end is constructed by using flexible braid to connect the copper plate to a movable copper block attached to the test section cluster. As a matter of interest, this assembly failed in burn out as shown in Fig. 10. In another similar assembly, shown in Fig. 11, incipient burn out was detected and the power removed before complete burn out occurred.

Another type of test section utilizes the same cylindrical-type connection as the rectangular channel, except it is considerably larger for higher currents. The section itself is machined so that the

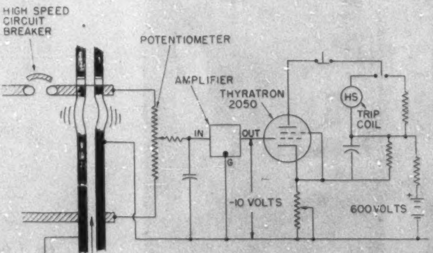


Fig. 12. Burn-out detector

heat is generated in a cosine distribution along the length with the maximum generation at the middle.

Still another section utilizes another terminal type. In this test section, an assembly of plates similar to the one described by Cunningham and Boyle[2] is used (see Fig. 2). The assembly was constructed of zirconium alloy and terminals had to be attached in such a way that the individual plates would be evenly heated without burning out at the terminals.

A connection in which two heavy copper terminal pieces were butt-joined to the edge of the zirconium assembly was tried. These terminals were large so that they would not deform under bolt stress. Of course, the bolt stress itself could not be high enough to deform the zirconium assembly. The length of contact on the sides was made about the same as the width of the plates. Thus the current flowing in would not tend to concentrate and exceed current densities farther down the plates.

The heavy contact surfaces were water-cooled through holes drilled in the copper just beneath the surface. This cooling of the copper close to the point of contact was provided in the hope that the heat generated in the contact and diffusion section of zirconium would be quickly dissipated. The results of the tests confirmed this cooling method. Although the total current through these terminals was raised gradually over a period of an hour from 6,000 to 15,500 amp and burn out occurred, there was no evidence of burning under the contacts. The conducting area of contact was about 9 square inches, carrying about 1,600 amp per square inch if the total was distributed uniformly over the contact surfaces.

In later tests shorter plate assemblies were used. Rather than change the terminal type, a shorter contact surface was used. In one test, soldered shims were also inserted. An area about half the original was used, and, with this combination, burning at the contacts did occur, although the main burn out still took place between the terminals.

Burn-Out Detection

As might be expected, the various test sections are expensive even though relatively inexpensive compared with reactors. For this reason, it is desirable to obtain burn-out data without destroying the test section. If the start of burn out can be detected, the power can be interrupted before the test section actually reaches a damaging temperature of 1,500 to 1,600 F. The most difficult type of burn out to detect is the fast burn out occurring at high-heat fluxes and high subcooling.

Before describing the means for detecting burn out it may be well to review the mechanism of burn out. Fig. 1 shows that heat transfer to water passes through several phases. The nucleate- and film-boiling phases are those important in burn out. As power is increased, a condition is reached where the heat-transfer action in the water changes from nucleate boiling to film boiling. As can be seen, the wall temperature for film boiling at the same power rate is in the failure region for the test section metal.

The rate at which burn out proceeds can be expected to be a function of both the thermal capacity of the metal and the difference between the heat generated and that removed. This difference is the space between the horizontal line and the curve. A more conservative estimate would be based on the assumption that all the heat goes into the metal.[4]

As the temperature of the metal rises, the yield point decreases. Failure of the metal is assumed to take place when the yield-point strength drops to the point where it just equals the applied stress. Test sections have been used or considered where the burn-out time, as computed in the foregoing varies from about 0.010 second to nearly a second or more. In order to detect the fast burn out and prevent test section failure, the detection and power-interruption system must be very fast.

Because the time delay caused by the d-c main current circuit breaker could be the major problem, a Westinghouse high-speed breaker was modified to provide a shunt-trip bucking coil on the holding-coil frame. This coil was specially wound with a low-inductance coil of fewer than normal turns. Since the tripping time was expected to be less than 1/2 cycle, the charge on a capacitor could be utilized to energize this coil and to trip the breaker.

The whole tripping circuit as finally developed is shown in Fig. 12. It consists of a temperature excursion detector, a high-gain amplifier, a thyratron tube, the capacitor circuit, and the high-speed breaker. The amplifier raises the signal from millivolts to volts and feeds this signal into the thyratron, causing it to fire at a desired signal level. The thyratron permits the capacitor to discharge through the trip coil of the high-speed breaker. The interruption time has been measured to be 0.005 second from detection of signal to interruption of a 1,000-amp current. This includes the delay time in the detector and in the high-speed breaker.

The temperature excursion is detected either by thermocouple or by a resistance change in the test section. The resistance change is a function of the actual temperature and is used where high-speed detection is needed. Burn out occurs near the end of the test section where the water is hottest. This area heats much more rapidly than the rest of the specimen. The resistance of the section is, of course, a function of temperature. The voltage of a centrally located lead on the test section moves toward that of one terminal when burn out starts to occur. By balancing the center point of a potentiometer against the voltage of this lead and adjusting the voltage difference to zero prior to burn out, it is possible to detect this voltage shift and to use it to detect burn out; see Fig. 12.

For slower speed burn outs, the thermocouple type of detection can be used. It can be made to operate a relay and to trip the main breaker or the generator field breaker as desired or required. Paralleling of thermocouple and resistance bridge through separate amplifiers to cause power interruption by either means is also possible. A scheme has also been developed to operate several self-balancing bridges, so that burn out may be detected if it occurs in any one of several tubes or plates in parallel.

Where test sections are made from zirconium, resistance-bridge detectors do not seem to be very reliable because the characteristic resistance curve of the metal takes a dip in the region of the burn out temperature.[5] Fortunately, all tests with this metal have been made on specimens of large wall thickness with slow-speed burn out; consequently, thermocouples attached through thin insulation have apparently been fast enough. If higher speeds are required, thermocouples attached directly to the test section may be used although there may be some calibration problems due to the high currents passing the tips of the couple.

Summary

Heat-transfer burn out is important in determining the maximum permissible power level of nuclear reactors. This paper describes some of the electrical problems encountered in the design and operation of an electrical heat-transfer burn-out test facility used to determine these maximum permissible power levels. The facility currently includes a 300-kw power supply for heating the test sections. There is an additional 2,500-

kw power supply which is being installed.

A number of unusual problems were solved in connection with conducting the high electric currents into test sections where current densities as large as 173,000 amp per square inch were required. The test sections were cooled by 2,000 psig and 630 F water. Several methods for detecting incipient burn out under these conditions were developed.

References

1. Heat Transmission (book), **W. H. McAdams.** McGraw-Hill Book Company, Inc., New York, N. Y., third edition, 1954, pp. 370, 389–90.

2. MTR-Type Fuel Elements, **J. E. Cunningham, E. J. Boyle.** *Proceedings*, International Conference on the Peaceful Uses of Atomic Energy, vol. 9, "Reactor Technology and Chemical Processing," 1956, pp. 203-07.

3. Burnout Heat Flux in Rectangular Channels, **D. A. Dingee, H. M. Epstein, J. W. Chastain, S. L. Fawcett.** *BMI 1065*, Battelle Memorial Institute, Columbus, Ohio, Jan. 10, 1956.

4. Boiling Studies, **A. G. Guibert, F. E. Romie.** *COO-24*, Department of Engineering, University of California, Berkeley, Calif., Aug. 1950.

5. Pitting Anodic Corrosion Studies, **H. S. Jacket.** *WAPD-RcM-16*, Westinghouse Atomic Power Division, Bettis Plant, Ptitsburgh, Pa., Nov. 6, 1953.

6. Metallurgy of Zirconium (book), **B. Lustman, F. Kerze,** editors. McGraw-Hill Book Company, Inc., New York, N. Y., 1955, "Nationa Nuclear Energy Series, Division VII," vol. 4, p 366, Fig. 8.5.

Acoustic Ionization Detector

DAVID R. WHITEHOUSE
NONMEMBER AIEE

FRANK S. REPLOGLE, JR.
NONMEMBER AIEE

Synopsis: A unique a-c ionization chamber is described utilizing the acoustic pressure generated within an ionized gas by an applied a-c field. On the basis of preliminary theory an experimental neutron-sensitive chamber was designed. The instrument has an approximate sensitivity of 1.5×10^7 neutrons/cm²/sec (neutrons per square centimeter per second) and proves to be rugged and fast.

THE observance of events on the atomic or subatomic level is achieved at present by ionization chambers, counters, cloud chambers, electroscopes, scintillation detectors, bolometers, crystal counters, photograph emulsions, and other more specialized types of detectors. The need for these different detectors has arisen because of the many types, intensities, and energies of the presently encountered radiations. Radiation detectors are widely used in many branches of research and industry. They may be used to establish the identity and amount of naturally radioactive material or artificially activated substances. They are singularly important in cosmic-radiation studies, and an important technique of coincident counting has evolved. It is now possible to analyse very complex ionizing events. These detectors also find extensive use in medicine, industrial tracers, petroleum geology, chemical-reaction rates, nuclear instrumentation, etc. This paper is primarily concerned in the development of a detector for nuclear-reactor instrumentation.

Nuclear reactors have presented the peculiar problem of measuring neutron flux over many decades of range. This fact precludes the use of only one type of detection channel and several instruments must be used to cover the complete range. For low-level measurements, the Geiger or proportional chamber is used with the associated pulse-counting equipment. For high-level measurement, an ionization chamber is used with an integrating circuit so that the pulses are smoothed, and the average current is measured. All of these chambers are operated with d-c collection, or polarizing potentials. The limit of sensitivity is often determined by the drift in the d-c amplifiers, and the response time is slow from the long *R-C* time constant of the detection circuit. Likewise, the circuit which polarizes the ion chamber is used to measure the signal current. Therefore the insulators used in ionization chambers must have a resistance greater than 10^{13} ohm for efficient operation, and this specification is not easily met since these chambers must function in high-flux regions where radiation damage is a serious problem.

A-c operation of ionization chambers has been proposed to eliminate the problem of drift in d-c systems. It is possible to use the chamber in an a-c bridge circuit where the chamber acts as a leaky capacitor. The unbalance of the bridge is then proportional to the incident radiation. A report by A. H. Ward[1] describes a system wherein two ion chambers are used in the bridge, and a sensitivity of 10^{-14} amperes is obtained. The time constant of this instrument, however, is about 15 sec. If the sensitivity is not too important, fast chambers can also be realized. The problem of insulation destruction, though somewhat relaxed, is still an important consideration.

A research study has been performed on a novel type of ionization chamber where, in addition to a-c operation of the chamber, the output signal has been completely isolated from the polarizing electrodes. The chamber converts electric energy into acoustical energy, and the signal is measured as an acoustical pressure.

Principle of Acoustic Ionization Chamber

It is observed that for ionization chambers with moderately high gas pressures, the electric energy which is dissipated in the chamber is spent in the gas rather than in the electrode. The implication of this statement is that a possible method of radiation detection would be to measure the energy increase of the gas rather than to measure the current that flows. By using an audio-frequency polarizing voltage, the increase in energy of the gas will occur in pulses, and a sound wave will be generated in the chamber. The measurement of the sound intensity in this a-c acoustic ionization chamber is proportional to the ion density which in turn is proportional to the incident radiation.

The principle of operation is thus directly concerned with events on the atomic scale which point to the generation of sound by ionic motion. It is a fact that acoustic pressure waves in a gas are a result of particle motion. These particles or molecules move in the direction of propagation of the wave and give rise to alternate compressions and rarefactions in the gas. One of the most common methods of producing these varia-

---

Paper 57-408, recommended by the AIEE Nucleonic and Radiation Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE 2nd Nuclear Engineering and Science Conference, Philadelphia, Pa., March 11–14, 1957. Manuscript submitted March 11, 1957; made available for printing June 19, 1957.

David R. Whitehouse is with Massachusetts Institute of Technology, Cambridge, Mass., and Frank S. Replogle, Jr. is with the Schlumberger Well Surveying Corporation, Ridgefield, Conn.

This study was made possible through support extended the Electronic Nuclear Instrumentation Group at the Servomechanisms Laboratory, Department of Electrical Engineering, Massachusetts Institute of Technology, by the Office of Naval Research and the Navy Bureau of Ships under contract *N5ori-07876*. The authors express their appreciation to Prof. T. S. Gray for his constructive suggestions and supervision of the project.

tions is to excite a piston or diaphragm which is emersed in the gas. The layer of gas next to the piston is then alternately compressed and expanded, and the tansfer of energy to the gas occurs almost adiabatically.

It is logical therefore that if the molecules of a given volume of gas have their energy uniformly increased, a pressure or sound wave will be generated into the surrounding volume of gas. Since a positive ion is a molecule with an electrical "handle" on its charge, motion can be imparted to this ion when it is in the region of an electric field. As the ion travels through the field, it receives an amount of energy which is proportional to its charge and the potential through which it has traveled. In gases at atmospheric pressure, the mean free path of an ion or molecule is in the order of 10^{-5} cm, and the energy which is gained by the ion will be quickly distributed to the other gas molecules. This thesis is proved in the following by looking at the conditions which prevail in the gas.

Assume that there is 1 cm^3 of a uniform gas, such as argon, in which there is an electrostatic field of 1,000 volts/cm. The following data,[3] given in centimeter-gram-second units, are then pertinent.

$\bar{v}=46\times10^{3}$ cm/sec = average molecular agitation velocity
$1/2(m\bar{v}^{2})=0.03$ ev (electron volt)/molecule = average molecular kinetic energy
$\lambda=6\times10^{-6}$ cm = mean free path of molecules
$\theta=(46\times10^{3})/(6\times10^{-6})=7.7\times10^{9}$ collisions/sec = molecular collision rate
$\mu=1$ cm^2/volt/sec = mobility of positive ions at 760-millimeters Hg
$W=1{,}000$ cm/sec = drift velocity of ion at 1,000 volts/cm
$T=1$ millisecond = travel time for a drift of 1 cm

The drift velocity of the ion is superposed on the agitation velocity which the ion would have if it were a molecule. This increase in velocity is only 2% of the agitation velocity and hence the kinetic energy of the ion while being collected is practically unchanged. Since the collisions limit the drift velocity, the 1,000 ev of energy which the ion obtains in traveling a distance of 1 cm is given almost entirely to the molecules by the mechanism of collisions which occur at 7.7×10^{9} collisions/cm.

The situation for electrons is somewhat different as shown by the following considerations. The electrons have a longer mean free path than ions and they lose much less energy per collision than the ions. Again, with the field at 1,000 volts/cm in a gas such as argon at atmospheric pressure, the data given in the following paragraph are pertinent.

$W=10^{6}$ cm/sec = drift velocity of electron
$E_a=10$ ev = agitation energy
$\lambda=10^{-4}$ cm = mean free path
$\bar{v}=2\times10^{8}$ cm/sec = rms agitation velocity
$\theta=2\times10^{8}/10^{-4}=2\times10^{12}$ collisions/sec = electronic collision rate
$T=1$ microsecond = travel time for a drift of 1 cm

The number of collisions in 1 cm is 2×10^{6} and the energy dissipated when the electron is collected at the electrode is the agitation energy which is 10 ev. It thus becomes evident that the efficiency of energy transfer from charged particles to molecules is less for electrons than for heavy ions. In fact, it is quite possible that in some cases the electrons will dissipate all of their received energy at the electrode and become completely inefficient at transferring energy to the gas.

From the classical kinetic theory of gases,[3] the pressure in a gas is equal to two thirds the total translatory kinetic energy of the molecules per unit volume. Since the relationship is linear, the incremental variation follows the same form.

$$\Delta P=2/3\Delta E_{\text{translatory}}$$

Thus one can obtain the rms pressure variations in a constant volume of gas by simply calculating the rms kinetic energy which is given to the molecules. It must be remembered, however, that some gases can store kinetic energy in the form of molecular rotation or vibration in addition to translation, and the energy which is added will divide itself equally among all the possible modes. Therefore, in the construction of an acoustic chamber, it would be wise to use a monatomic gas which has only three degrees of freedom, one for each direction of translation.

Adaptation to Neutron Detection

The acoustic ionization chamber is particularly adaptable to the detection of large thermal-neutron fluxes since they can excite rather large current densities within the chamber. With peak-current densities in the order of 1 microampere/cm^2, and peak field strengths in the order of 1,000 volts/cm, rms pressure variations in the order of 0.1 to 1.0 dyne/cm^2 can be generated at a frequency of 1,000 cps (cycles per second). If the chamber is acoustically resonated, the signal will be enhanced by an amount approximately equal to the resonant Q.

There are two possible standard techniques for detecting neutrons in an ionization chamber. One is to use a boron trifluoride gas, and the other is to use an elemental boron coating on the side walls of the chamber. The boron then reacts with the neutrons given by the following reaction:

$$B^{10}(n,\alpha)Li^{7}+Q(2.79 \text{ megelectron-volts})$$

The resultant energetic alpha and lithium particles copiously ionize the gas. Careful consideration however must be given to the ionization efficiencies of these particles. The efficiencies are determined by the particle range in the gas and by the geometry of the chamber. The ionizing alpha and lithium particles of the foregoing reactions will have particle ranges in gases at normal pressure of from 1 millimeter to 1 cm in length. With a given geometry, one can thus compute the approximate ionization efficiency by considering the percentage of the particle range which lies within the gas. The collection of these generated electrons and positive ions thus forms the primary current in the chamber and, if an a-c collection voltage is used, a pressure variation will be generated within the chamber.

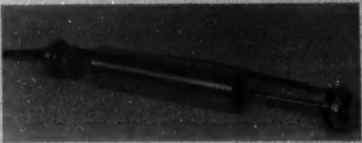


Fig. 1. A-c acoustic ionization chamber

The acoustic pressure may now be measured by a microphone placed inside the chamber. Since the microphone can have virtually any impedance level, a low-impedance microphone eliminates the problem of insulator destruction. The output is now obtained directly as an a-c signal. A new problem has been created, however, in that external noise may enter the chamber in the form of extraneous sound.

Construction of Neutron-Sensitive Chamber

The neutron-sensitive chamber was constructed for the purpose of demonstrating the previous ideas. Fig. 1 shows the completed chamber. The chamber is an aluminum cylinder 6 inches long with

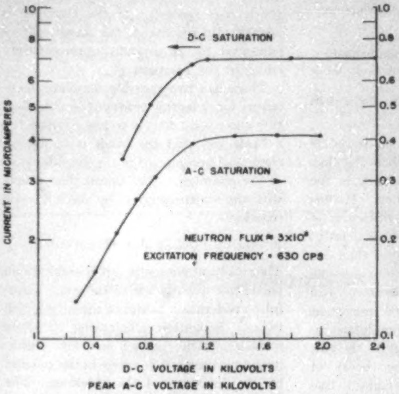


Fig. 2. Saturation curves for neutron-sensitive chamber

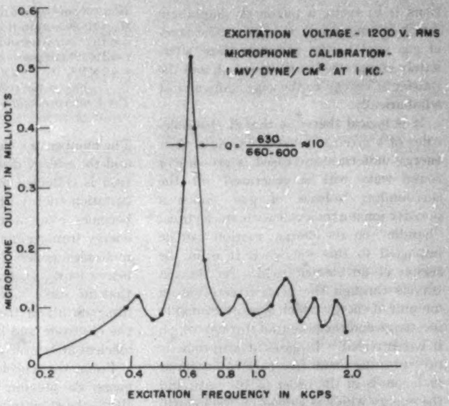


Fig. 3. Frequency response of neutron-sensitive chamber

1-inch inside diameter. An enriched-boron lining with an argon gas filling is used instead of the BF_3 gas for the following reasons:

1. The ionization can be localized in a certain volume.
2. The gas pressure can be varied without changing the number of neutrons which are captured.
3. Argon has three translatory modes of kinetic energy storage, whereas BF_3 likely has several more unuseful modes.
4. The chamber is much easier to fill with argon than with BF_3.

The principle of acoustic resonance is used in the chamber in order to increase its sensitivity. Resonance is accomplished by extending the boron lining and central electrode only through half the chamber and attaching the microphone to the opposite end. Resonance will occur when the length of the chamber corresponds to a half-wavelength of the frequency of the energy pulses.

Several microphones were investigated for use as the pressure-sensing device, among which were the crystal, carbon button, and magnetic elements. Since the microphone is also in the neutron flux, it was felt that the magnetic element would be most successful. A moving-armature type of magnetic microphone was obtained in the shape of a cylindrical disc with 1-inch diameter and 0.3-inch thickness. The sensitivity of the microphone was measured to be 1 millivolt/dyne/cm² at 1 kc, and the frequency response was nominally flat from 1 to 3 kc.

The thickness of the enriched-boron lining was approximately 1 milligram/cm.[3] The lining was placed on the inside of a thin cylindrical graphite shell which was then telescoped inside the aluminum tube. For the given dimensions, the gas pressure was set at 40 cm of Hg in order to obtain a uniform ion generation, and still maintain a good ionization efficiency. With the chamber operating at 1,200 volts, the current becomes space-charge limited for a neutron flux greater than 3×10^9 neutrons/cm²/sec.

When the alternating voltage is applied to the chamber, the input power will be the second harmonic of the excitation frequency. The chamber will then resonate at this second harmonic. Thus the excitation frequency was calculated to be approximately 600 cps. The average collection time of the positive ion for a voltage of 1,200 volts is approximately a half-period of 4,000 cps. The ion collection time, therefore, does not limit the frequency response of the output pressure.

Experimental Results

SATURATION CURVES

The chamber was tested at the reactor of the Brookhaven National Laboratory. Fig. 2 shows a plot of the d-c and a-c saturation curves for a thermal-neutron flux level of approximately 3×10^9 neutrons/cm²/sec. The d-c saturation curve is a plot of the primary electrode current as a function of the d-c polarizing voltage, and the curve shows a rather marked saturation voltage at 1.2 kv. The a-c saturation curve is a plot of the microphone output as a function of the a-c peak voltage at the resonant frequency of 630 cps. Since the acoustic pressure is also proportional to the amplitude of the excitation voltage, the microphone output is divided by the excitation voltage so that the saturation curve will be horizontal when all the ions are being collected. It is clearly evident that the peak a-c saturation voltage corresponds very closely to the d-c saturation voltage.

ACOUSTIC OUTPUT

The voltage was then set at 1,200 volts rms and the frequency was varied from 200 to 2,000 cps. The output of the microphone is shown in Fig. 3 with the resonant frequencies very pronounced. It must be remembered that the output of the microphone, in most cases, is twice the frequency of the excitation voltage, since each half of the sine-wave excitation will give an energy pulse to the gas. The largest output occurs for a driving frequency of 630 cps which corresponds to a chamber resonated at a half-wavelength of 1,260 cps, the output frequency. The acoustic Q of the chamber is determined by the frequency response which gives

$$Q=\frac{\omega_0}{\Delta\omega}=\frac{630}{660-600}\approx 10$$

This value corresponds very closely to measurements taken on the chamber with a speaker-and-microphone technique.

ENERGY CONSIDERATIONS

If the acoustic source, or energy input, were distributed longitudinally in the chamber as a sinusoid, the output would be approximately Q times the amplitude of the sinusoid. The source, however, is actually distributed as a nonsymmetrical square wave. Therefore, neglecting the higher harmonics, Q times the fundamental of the square-wave energy distribution gives a very approximate calculation for the output. The fundamental is $2/\pi$ times the amplitude of the nonsymmetrical square wave.

Thus with a Q of 10, the output at resonance, which is 0.52 millivolt, corresponds to a primary pressure generation of

$$\Delta P = \frac{0.52}{1 \times 10 \times 2/\pi} = 0.08 \text{ dyne/cm}^2$$

This value of pressure generation is now compared with the calculation of pressure variation from energy considerations. Assuming that the collection of the ion current is sinusoidal with a peak value equal to the d-c saturation value is in phase with the collection voltage, the instantaneous power is given by

$$P(t) = v(t)i(t) = \sqrt{2}(1{,}200) \sin \omega t \times (7 \times 10^{-9}) \sin \omega t$$

$$P(t) = 0.0119 \sin^2 \omega t = \frac{0.0119}{2}(1 - \cos 2\omega t)$$

and the energy input is given by

$$E(t) = \int P(t)dt = 0.0059t - \frac{0.0059}{2\omega} \sin 2\omega t$$

The rms value of the energy variation at 1,260 cps is then

$$\Delta E = \frac{0.0059}{\sqrt{2} \times 2\pi \times 1{,}260} = 5.28 \times 10^{-7} \text{ joule} = 5.28 \text{ ergs}$$

With a volume of 25 cm³, and assuming that all the energy goes into translatory kinetic energy of the molecules, the pressure variation is

$$\Delta P = \frac{2}{3} \Delta E/\text{volume} = \frac{2}{3} \times \frac{5.28}{25} = 0.14 \text{ dyne/cm}^2$$

This value corresponds very closely to twice the measured value of 0.08 dyne/cm². This efficiency of 50% is easily accounted for by considering that the electrons, which receive 50% of the input power, do not transfer their energy to the kinetic energy of the gas, but dissipate it in exciting the gas atoms and heating the collection electrode. Thus the foregoing calculation is a good check on the theories which have been proposed.

TIME RESPONSE

The time response of the chamber was measured by applying step function changes in the amplitude of the a-c excitation. This method is not exactly equivalent to a step change in the neutron flux. However, the result gives an indication of the time necessary for the re-establishment of equilibrium. The time to reach 63% of the new level was 3 to 5 cycles of 1,260 cps, which is approximately 2 to 4 milliseconds. This time delay, of course, is due mainly to the resonance of the chamber and not the ionization effects. To show this, the time response was measured at off-resonant frequencies and the output was seen to respond within one cycle of the operating frequency. Hence 1-cycle response could be obtained if the chamber were designed to operate at a nonresonant frequency.

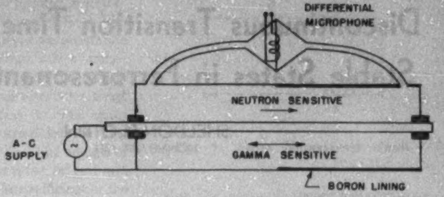


Fig. 4. Proposed neutron - sensitive chamber with compensation against gamma-induced ionization

DYNAMIC RANGE

The dynamic range of this instrument is set on the low end by inherent noise and on the high end by saturation effects. Over a frequency band of 400 to 5,000 cps, the noise level was 0.025 millivolt. For this measurement the chamber was excited with 1,200 volts at 630 cps and placed in the reactor with a gamma background but no neutron flux. Since the maximum output is 0.52 millivolt, the dynamic range is 20 to 1. This is by no means the limit, however, because the chamber can be designed to accommodate a larger neutron flux without affecting the noise limitation on the low end. An increase in gas pressure and smaller electrode displacement with higher field strengths should increase the dynamic range to at least 2,000 to 1.

Summary

The test of the neutron-sensitive chamber indicated that it can be useful as a control instrument. In addition, it seems logical that with further experiments, this technique of sound generation can be used to investigate physical events on the molecular scale.

As an instrument, the chamber has the following characteristics:

1. Minimum detectable flux level of 1.5×10^7 neutrons/cm²/sec. This flux level corresponds to the noise of the chamber as measured by the microphone. This noise may be either acoustical noise or gamma background noise. To minimize the latter effect, the chamber could be designed as in Fig. 4 where the entire chamber is gamma-sensitive, but only half the chamber is neutron-sensitive. A differential microphone then measures the pressure difference between the two ends of the chamber. This pressure difference is proportional only to the neutron flux.
2. Dynamic range of 20 to 1 (measured), 2,000 to 1 (estimated readily attained). The dynamic range is defined here as the ratio of the saturation signal level to the noise level. The estimated range of 2,000 to 1 is obtained by increasing the saturation signal.
3. Time response of 2 to 4 milliseconds. This time response is limited by the resonant Q of the chamber and is adequate for the control of most reactors.

In addition, the chamber is essentially a very rugged instrument. The reliability and life of the chamber are hard to determine at this time. The most sensitive part of the device is the pressure transducer which must withstand conditions inside a nuclear reactor.

The instrument requires a medium-high alternating voltage for excitation at a power level of less than 1 watt. Such a source could possibly be the 400-cycle supply which many laboratories maintain. In addition to the power cable, the chamber requires a second cable for the output of the microphone. The impedance of the output may be at almost any level, depending on the type of microphone used, and as a result the restrictions on the insulator resistance have been completely relaxed.

References

1. LOW FREQUENCY OPERATION OF BALANCED ION CHAMBERS IN AN A-C BRIDGE, A. H. Ward. *AERE-EL/R-1107*, Atomic Energy Research Establishment, Harwell, England, Jan. 27, 1953, p. 11.

2. IONIZATION CHAMBERS AND COUNTERS (book), B. B. Rossi, H. H. Staub. McGraw-Hill Book Company, Inc., New York, N. Y., 1949, chap. 1.

3. KINETIC THEORY OF GASES (book), E. H. Kennard. *Ibid.*, 1938, chap. 1.

Discontinuous Transition Time Between Stable States in Ferroresonant Circuits

SHELDON PLOTKIN
NONMEMBER AIEE

Synopsis: The nonlinear differential equations for series and parallel ferroresonant circuits are solved to give the discontinuous transition time between stable states. This time agrees with experiment. Such things as the dependence of "jump time" on lamination thickness and a discussion of some general negative-resistance factors are given.

The mathematical solution is accomplished by assuming the saturation curve of the inductor to be a series of odd-powered functions and the investigation is limited to the cubic term. First, a single sinusoidal solution is assumed and then a second term, third-harmonic correction, is included. The static results are compared with the actual observed circuit characteristic as well as with a simple graphical solution. Transient analysis is carried out using a variational approach which allows the perturbed differential equations to be reduced to canonical form. This results in a damping-time constant and oscillation frequency both of which are nonlinear. Hysteresis and eddy-current losses are found to have a significant effect and are included in the study.

IN RECENT years an increasing effort has been expended toward the solution of hitherto unsolved nonlinear problems. At the same time industry has developed a variety of new and better magnetic core materials so that in the future it is expected that ferroresonant systems will find use in an increasing number of applications. At present, some applications using the ferroresonant principle are magnetic amplifiers, sensing devices, stabilizers, and counting circuits.

With the inherent analytical difficulties involved with solutions to nonlinear problems, the number of methods of attack and useful approximation techniques is never large enough. It is hoped that this study will add something to the readers' insight and understanding of nonlinear problems in general and, specifically, will develop a better feeling for the actions of series and parallel ferroresonant circuits under both static and transient conditions. An analytical method for determination of the transient response is given following Liapounoff's approach to the investigation of stability in nonlinear oscillatory systems. The method is quite general and could be applied in a large number of problems. From the analysis, the nonlinear transition speed as well as a detailed derivation of the static circuit characteristic for both the series and parallel cases are found to be significantly affected by the hysteresis and eddy-current losses. It is these losses that contribute, in the main, to the rather wide difference often found between experimental and theoretical results. Agreement in the present study is excellent. Thus a good indication is given of the degree of approximation necessary to obtain a given degree of agreement.

Nomenclature

A = peak amplitude of fundamental component of assumed flux
a = peak amplitude of assumed solution
a_0 = amplitude of steady-state solution
a_1 = coefficient in general first-order system of differential equations
a_2 = coefficient in general first-order system of differential equations
α = symbol used in differential equation for series circuit
α_1 = symbol used in differential equation for parallel circuit
B = peak amplitude of assumed third-harmonic component of flux
b_1 = coefficient in general first-order system of differential equation
b_2 = coefficient in general first-order system of differential equation
β = symbol used in differential equation for series circuit
β_1 = symbol used in differential equation for parallel circuit
C = circuit capacitance
c = symbol used in differential equation for series circuit
c_1 = symbol used in differential equation for parallel circuit
γ = saturation parameter of magnetic core
η = transformation variable
η_0 = initial value of η
f_D = damping frequency, cycles per second
K = constant
i = instantaneous value of total current
i_{rms} = rms value of total current
i_L = inductor current
i_c = condenser current
L_0 = inductance at zero current
ω = forcing frequency, radians per second
ω_D = damping frequency, radians per second
ω_0 = natural resonant frequency at zero current, series circuit
ω_{01} = symbol used in differential equation of parallel circuit
ϕ = magnetic flux
ϕ_1 = steady-state value of flux
ϕ_{rms} = rms value of ϕ
R = series circuit resistance
R_{core} = equivalent effective series resistance of inductor
R_h = hysteresis component of R_{core}
R_1 = driving-source impedance in parallel circuit
R_0 = external resistance in series with inductor
R_m = maximum theoretical resistance in series with inductor
R_{Lm} = same as R_m but applied to parallel circuit
S_1 = root of the characteristic equation
S_2 = root of the characteristic equation
τ = damping-time constant of perturbation
T = time
t = lamination thickness
V_m = peak voltage
v_1 = perturbation
v_2 = time derivative of v_1
W = peak voltage of sine component of forcing voltage
ξ = transformation variable
ξ_0 = initial value of ξ
Y = peak voltage of cosine component of forcing voltage

Theoretical Analysis, Series Ferroresonant Case

STATIC-CIRCUIT CHARACTERISTIC

The circuit to be analyzed first is that given in Fig. 1. The eddy-current and hysteresis losses R_{core} can be approximately represented as being included in the resistance R, where

$$R = R_{core} + R_0$$

The resistance R_{core} also includes the d-c winding resistance of the inductor. Any external resistance placed in the circuit is given by R_0.

According to Kirchoff's voltage law, the circuit equation is

$$\frac{d\phi}{dT} + Ri + \frac{1}{C}\int i\,dT = V(T) = V_m \sin\,(\omega T + \theta) \tag{1}$$

Differentiating equation 1 gives

$$\frac{d^2\phi}{dT^2} + R\frac{di}{dT} + \frac{i}{C} = \omega V_m \cos\,(\omega T + \theta) \tag{2}$$

The problem at hand lends itself well to assuming that the saturation of the inductor is cubic in nature; therefore, the odd powers above the third have been neglected. Additional powers of ϕ could be used but the additional labor involved

---

Paper 56-789, recommended by the AIEE Solid State Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 26, 1956; made available for printing February 7, 1957.

SHELDON PLOTKIN is with the University of California, Berkeley, Calif.

Sincere appreciation is extended to the following: Prof. R. S. Mackay, under whose supervision this work was done, for his invaluable assistance and criticism; Prof. S. P. Diliberto for his mathematical advice; and Professors J. R. Woodyard and W. S. Pritchett for their helpful suggestions.

with their inclusion was considered prohibitive at this stage of analysis.

$$i=\frac{\phi}{L_0}+\gamma\phi^3 \qquad (3)$$

L_0 represents the initial inductance of the core, whereas γ represents the saturation effect.

Differentiating equation 3 and substituting into equation 2 obtains

$$\frac{d^2\phi}{dT^2}+c\frac{d\phi}{dT}+\omega_0^2\phi=Y\cos\omega T-W\sin\omega T+\alpha\phi^2\frac{d\phi}{dT}+\beta\phi^3 \qquad (4)$$

where

$Y=\omega\ V_m\cos\theta$
$W=\omega\ V_m\sin\theta$
$\alpha=-3\gamma R$
$\beta=-\gamma/C$
$c=R/L_0$
$\omega_0^2=1/L_0C$

The reason for including a phase angle in the forcing term of the differential equation is to allow representing the solution without a phase angle.[1] Because the damping term contains only positive terms, it is impossible for the circuit to exhibit self-excited oscillations.

A solution to the differential equation is assumed to be $\phi=A\ \cos\omega T$.[1,2] Substituting this into equation 4 and equating sine and cosine terms, neglecting higher harmonics and higher order terms, gives

$$\left.\begin{aligned}c\omega A-\alpha\omega A^3&=W\\(\omega_0^2-\omega^2)A-\frac{3\beta A^3}{4}&=Y\end{aligned}\right\} \qquad (5)$$

Squaring and adding equation 5 yields

$$\omega^2V_m^2=A^6\left[\frac{1}{16}(\alpha^2\omega^2+9\beta^2)\right]+A^4\left[(\omega^2-\omega_0^2)\frac{3\beta}{2}-\frac{c\alpha\omega^2}{2}\right]+A^2[(\omega_0^2-\omega^2)^2+c\omega^2] \qquad (6)$$

By relating equation 6 to equation 3 the static-circuit characteristic is determined, see Table I. In order for a discontinuity to exist a negative resistance must prevail, and because i is a monotonically increasing function of ϕ, the negative-resistance characteristic must be contained in equation 6. The exact points of discontinuity are determined by setting $d(V_m^2)/d(A^2)=0$. The square of the flux at these points must be real and positive. Therefore, the coefficient of the A^4 term in equation 6 must be negative.

Table I. Series Static-Circuit Characteristic for $\phi=A\ \text{Cos}\ \omega T$

$$V_{rms}^2=\phi_{rms}^6\left[\frac{9}{4}\gamma^2\left(R^2+\frac{1}{\omega^2C^2}\right)\right]+\phi_{rms}^4\left[3\gamma\left(\frac{R^2}{L_0}-\frac{\omega^2-\omega_0^2}{C\omega^2}\right)\right]+\phi_{rms}^2\left[\frac{(\omega_0^2-\omega^2)^2}{\omega^2}+\left(\frac{R}{L_0}\right)^2\right]$$

ϕ RMS, Volt-Second	R, Ohms	i RMS, Ampere	V^2 RMS, Volts^2	V RMS, Volts
0.1	3,000	0.001	1,191	35
0.2	1,000	0.010	2,453	50
0.30	500	0.035	540	23
0.4	0	0.082	4,400	67

$$(\omega^2-\omega_0^2)\frac{3\beta}{2}-\frac{c\alpha\omega^2}{2}<0 \qquad (7)$$

After substituting the values of α and β from equation 4, one obtains

$$\frac{CR^2}{L_0}<1-\frac{\omega_0^2}{\omega^2} \qquad (8)$$

This inequality shows that a condenser must be present in the circuit; CR^2/L_0 cannot be too large, and ω_0 must be less than ω. Recalling the value of ω_0, it is concluded that $\omega L_0<1/\omega C$. Thus the circuit is initially inductive, and it is reasoned that as the reactor saturates the circuit must eventually become capacitive.

Note that if C approaches infinity, i.e., no condenser in the circuit, the circuit equation from equation 6 becomes

$$\omega^2V_m^2=A^6\left[\frac{1}{16}\alpha^2\omega^2\right]+A^4\left[\frac{3c\gamma R\omega^2}{2}\right]+A^2[\omega^4+c^2\omega^2] \qquad (9)$$

For bistability to exist with no capacitor, γ must be less than zero, which indicates a negative slope to the saturation curve (a physically unrealizable situation at present).

Stipulating that the coefficient of A^4 be negative is only a necessary condition. A further sufficient condition is that the discriminate, when solving for A^2 from the $d(V_m^2)/d(A^2)=0$ equation, be positive.

$$[(\omega^2-\omega_0^2)3\beta-c\alpha\omega^2]^2-\frac{3}{4}(\alpha^2\omega^2+9\beta^2)\times[(\omega_0^2-\omega^2)^2+c^2\omega^2]>0 \qquad (10)$$

If it is assumed that $\omega L_0>>R$ and that $\omega>>\omega_0$, which is in good agreement with experimental results, the maximum possible theoretical resistance may be determined.

$$R_m\approx\frac{1}{\sqrt{3}\omega C} \qquad (11)$$

At this value for resistance, the negative resistance part of the over-all circuit characteristic disappears and no discontinuity is possible. It follows that this result is only as good as the assumptions for L_0 and ω_0. As ω_0 increases in value, the maximum theoretical resistance decreases. From the experimental results, equation 11 is only within 15%. A closer approximation is indicated in the analysis of the parallel circuit.

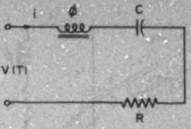


Fig. 1. Series ferroresonant circuit

Considering the static-circuit impedance at different points, it is known that initially it is inductive. As the voltage and current increase, the inductor begins to saturate. Thus the impedance of the circuit decreases with rising current until a minimum impedance point is reached. At currents above this point the impedance increases. Because the saturation curve does not possess a negative slope, it can be concluded that the minimum impedance point occurs when the inductive reactance is cancelled by the capacitive reactance, at least with regards to the fundamental component which is dominant.[3] Thus the increasing over-all circuit impedance above the minimum point must be capacitive because if it were not, the over-all impedance could not continue to increase. It follows then that the point of minimum over-all circuit impedance is the resonant point and the equivalent series resistance (R_0+R_{core}) of the circuit can be calculated there from the static-circuit characteristic.

Equation 6 assumed a flux consisting of only a fundamental term. The analysis is now extended to include the third harmonic of the flux or second approximation.

$$\phi=A\ \cos\omega T+B\ \cos3\omega T \qquad (12)$$

Table II. Series Static-Circuit Characteristic for $\phi=A\ \text{Cos}\ \omega T-A/6\ (\text{Cos}\ 3\omega T)$

$$V_{rms}^2=\phi_{rms}^6\left[\frac{16}{9}\gamma^2\left(R^2+\frac{1}{\omega^2C^2}\right)\right]+\phi_{rms}^4\left[\frac{8}{3}\gamma\left(\frac{R^2}{L_0}-\frac{\omega^2-\omega_0^2}{C\omega^2}\right)\right]+\phi_{rms}^2\left[\frac{(\omega_0^2-\omega^2)^2}{\omega^2}+\left(\frac{R}{L_0}\right)^2\right]$$

ϕ RMS, Volt-Second	R, Ohms	i RMS, Ampere	V^2 RMS, Volts^2	V RMS, Volts
0.1	3,000	0.001	1,214	35
0.2	1,000	0.010	2,720	52
0.3	500	0.035	1,170	34
0.4	0	0.082	2,500	50

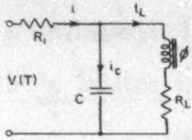


Fig. 2. Parallel ferroresonant circuit

Substituting this value of flux into equation 4 and equating coefficients of trigonometric terms, neglecting all harmonics higher than the third (see Appendix I), gives

$$\left.\begin{aligned} Y&=(\omega_0{}^2-\omega^2)A-\frac{3}{4}\beta A^3-\frac{3}{4}\beta A^2B-\\ &\qquad\frac{3}{2}\beta AB^2\\ W&=C\omega A-\frac{1}{4}\alpha\omega A^3-\frac{1}{4}\alpha\omega A^2B-\\ &\qquad\frac{1}{2}\alpha\omega AB^2\end{aligned}\right\}\qquad(13)$$

$$0=(9\omega^2-\omega_0{}^2)B+\frac{1}{4}\beta A^3+\frac{3}{2}\beta A^2B+\frac{3}{4}\beta B^3\qquad(14)$$

$$0=3cB-\frac{1}{4}\alpha A^3-\frac{3}{2}\alpha A^2B-\frac{3}{4}\alpha B^3\qquad(15)$$

It would be satisfying if equations 14 and 15 gave approximately the same result, but this is not the case. Equation 14 gives a positive value of B in terms of A while equation 15 gives a negative value. Considering the nature of the inductor saturation curve and the fact that the flux wave is considered to be cosinusoidal for this solution, the result of equation 14 is physically unrealizable. Saturable reactors are flux-saturating and at present a current-saturating reactor is unknown. Therefore, on physical grounds, the solution of equation 14 is considered to be unreal and is, therefore, ignored. The result of equation 15 is

$$B\approx-\frac{A}{6}\qquad(16)$$

This value is used throughout the remainder of the paper and is in good agreement with empirical evidence.

Using the second approximation for flux, equation 16, in equation 13, a slightly corrected static-circuit characteristic results (see Table II).

$$V_m{}^2=A^6\left[\frac{4\gamma^2}{9}\left(\frac{1}{\omega^2C^2}+R^2\right)\right]-A^4\left[\frac{4\gamma}{3}\left(\frac{\omega^2-\omega_0{}^2}{\omega^2C}-\frac{R^2}{L_0}\right)\right]+A^2\left[\frac{(\omega^2-\omega_0{}^2)^2}{\omega^2}+\left(\frac{R}{L_0}\right)^2\right]\qquad(17)$$

It is noted that in this approach when the assumed flux contains a single sinusoid a reasonable value of third-harmonic current results. However, when the third-harmonic component of assumed flux is included, an erroneous third-harmonic current component results and a good fifth harmonic prevails.

TRANSITION TIME

The method of analysis for discontinuous transition speed was used by Poincaré for classification of singularities and by Liapounoff for investigation of the stability of nonlinear systems.[4] This method has application in a large variety of second-order nonlinear problems whose differential equations can be written explicitly. Given a stable operating point, the system can be perturbed by a quantity v_1, and then after transformation to canonical form the damping of the perturbation can be determined. In this way both the damping-time constant as well as the damping frequency may be calculated for a given system.

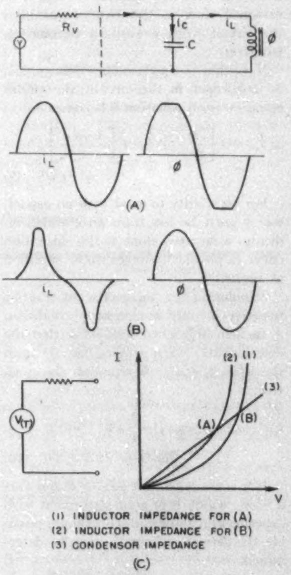


Fig. 3. Saturation waveforms and characteristics of reactor and capacitor

A—Waveforms using current-type source (large R_v)

B—Waveforms using voltage source ($R_v \approx 0$)

C—Rms volt-ampere characteristics of reactor and capacitor using high- and low-impedance sources

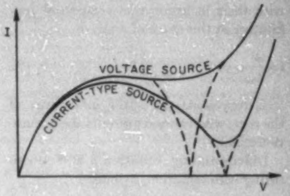


Fig. 4. Graphically derived parallel ferroresonant static characteristic taking into account resistance

Another method of solution is indicated by Hayashi, which integrates on a phase plane from one singularity to the next across the discontinuity.[5] This matter was not investigated. It necessitates a different approach to solution of the original differential equation.

Assuming $\phi=A\cos\omega T$, the system of differential equations for the perturbation can be written as follows

$$\phi=A\cos\omega T+v_1$$

$$\frac{d\phi}{dT}=-\omega A\sin\omega T+v_2\qquad(18)$$

From equation 4 the system of perturbation differential equations becomes

$$\frac{dv_1}{dT}=v_2$$

$$\frac{dv_2}{dT}=v_1\left(\frac{3\beta A^2}{2}-\omega_0{}^2\right)+v_2\left(\frac{\alpha A^2}{2}-c\right)+F_1(A,Y)\cos\omega T+F_2(A,W)\sin\omega T\qquad(19)$$

where

$$F_1(A,Y)=0$$
$$F_2(A,W)=0$$

$F_1(A,Y)$ and $F_2(A,W)$ are nothing more than the static-circuit characteristic, equation 5. All higher order and higher harmonic terms have been neglected.

Solution of the system, equation 19 is given in Appendix II* where

$$S^2-S(a_1+b_2)+(a_1b_2-a_2b_1)=0\qquad(20)$$

$$S=-\frac{1}{\tau}\pm j\omega_D\qquad(21)$$

where

τ = damping-time constant
ω_D = damping frequency

It is noted that the damping-time constant of the system is one half the rms value of the $d\phi/dT$ coefficient in the original differential equation 4. This fact might have been guessed from the outset.

---

* For further details, see reference 4, p. 43.

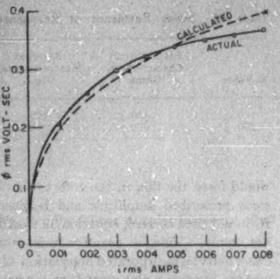


Fig. 5. Saturation curves for the primary winding of small filament transformer

The perturbations are given by

$$v_1=(S_2-a_1)\xi_0\epsilon^{S_1T}+(S_1-a_1)\eta_0\epsilon^{S_2T} \tag{22}$$

$$v_2=-a_2\xi_0\epsilon^{S_1T}-a_2\eta_0\epsilon^{S_2T} \tag{23}$$

Noting that S_1 and S_2 are complex conjugate, the imaginary terms of equation 22 cancel when ξ_0 and η_0 are equal. These quantities are not necessarily equal but their equality simplifies the mathematics without invalidating the general conclusion. If ξ_0 and η_0 are not equal, v_1 and v_2 will be out of phase. It is theoretically possible to use this method to a higher degree of approximation by including the higher order terms in the perturbation differential equations. However, this extension becomes exceedingly complex in that further nonlinear differential equations are involved with the circuit considered here.

Cancelling the imaginary components of equation 22 gives

$$v_1=\left(\frac{\alpha A^2}{2}-c\right)\xi_0\epsilon^{-T/\tau}\times\left(\frac{1}{\tau}\cos\omega_DT+\omega_D\sin\omega_DT\right)=v_1f(\omega_DT)\epsilon^{-T/\tau} \tag{24}$$

where

$$v_1=\xi_0\left(\frac{\alpha A^2}{2}-c\right)$$

$$f(\omega_DT)=\frac{1}{\tau}\cos\omega_DT+\omega_D\sin\omega_DT$$

The solution for the damping-time constant and damping frequency has been for the flux in the inductor. To determine the discontinuous current transition, one must relate the flux to current by equation 3.

$$\phi=\phi_1v_1f(\omega_DT)\epsilon^{-T/\tau} \tag{25}$$

$$i\approx\gamma\phi_1^2[\phi_1+3v_1f(\omega_DT)\epsilon^{-T/\tau}] \tag{26}$$

ϕ_1 is the flux in the inductor at the stable operating point after the jump. L_s has been assumed very large and additional time constants of $\tau/2$ and $\tau/3$ have been neglected.

The effect of the third-harmonic correction on the transition speed is now calculated. The same approach is followed as previously used except that the perturbation equations are

$$\left.\begin{aligned}\phi&=A\cos\omega T-\frac{A}{6}\cos3\omega T+v_1\\ \frac{d\phi}{dT}&=-\omega A\sin\omega T+\frac{A\omega}{2}\sin3\omega T+v_2\end{aligned}\right\} \tag{27}$$

Substituting equation 27 into equation 4, the system of perturbed differential equations are found to be

$$\frac{dv_1}{dT}=v_2$$

$$\frac{dv_2}{dT}=\left[-\omega_0^2+\beta A^2\left(\frac{3}{2}+\frac{1}{72}\right)\right]v_1+\left[-c+\alpha A^2\left(\frac{1}{2}+\frac{1}{72}\right)\right]v_2+F_{11}(A,Y)\cos\omega T+F_{22}(A,W)\sin\omega T \tag{28}$$

where

$$F_{11}=F_{22}=0$$

Comparing equation 28 with equation 19 shows that inclusion of the third harmonic in the transition time gives a completely negligible contribution and can therefore be ignored in the work to follow.

A second method of solution for transition time can be obtained from the method of Kryloff and Bogoliuboff.[4,6] The limitations of this method are the same as the variational approach so that only a check on the previous work can be expected. The limitations are twofold:

1. The amplitude of response is a slowly varying function of time compared with the period of the forcing function.
2. The nonlinearity is small.

These limitations are rather strained in the problem at hand. However, some insight into the adaptability of the methods used can be obtained from the experimental results. The Kryloff-Bogoliuboff method treats a differential equation of the form

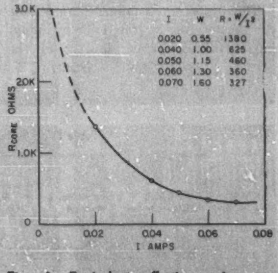


Fig. 6. Equivalent effective series core resistance

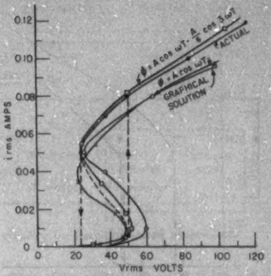


Fig. 7. Static-circuit characteristic curve for the series ferroresonant case; C = 1.0 μf, R_s = 0

$$\frac{d^2x}{dT^2}+\omega_0^2+\mu f\left(x,\frac{dx}{dT},T\right)=0 \tag{29}$$

Assuming $x=a\cos(\omega T+\theta_1)$ is a solution to equation 29, the following result

$$\frac{da}{dT}=\frac{-\mu}{2\pi a}\int_0^{2\pi}f(a\cos\omega T,-a\omega\sin\omega T,T)\sin\omega T\,d(\omega T) \tag{30}$$

$$\frac{d\theta_1}{dT}=\frac{\mu}{2\pi\omega a}\int_0^{2\pi}f(a\cos\omega T,-a\omega\sin\omega T,T)\cos\omega T\,d(\omega T) \tag{31}$$

Relating equation 29 to equation 4, setting $W=0$ because the phase angle is now included in the solution, and carrying out equation 30, gives

$$\frac{da}{dT}=\frac{3\gamma Ra^3}{8} \tag{32}$$

$$T=\frac{4}{3\gamma Ra^2} \tag{33}$$

In this problem both R and a are nonlinear so that the transition speed is constantly changing during the discontinuity. If equation 33 is considered to be an approximation to an exponential build-up, there results

$$a=a_0(1-\epsilon^{-T/\tau})\approx a_0[1-(1-T/\tau)]=T\frac{a_0}{\tau} \tag{34}$$

where τ is the time constant of the system and equals $4/3\gamma Ra^2$.

The change in phase θ_1 is rather difficult to relate to ω_D of the previous calculation and therefore that integration and comparison were not carried out. In this case

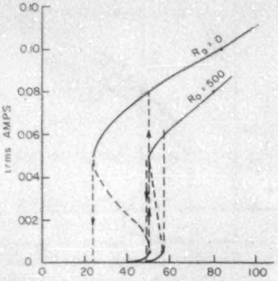


Fig. 8. Experimental static series-ferroresonant-circuit characteristic with variable external resistance

the difference in the calculated damping-time constant of the two methods is a factor R/L_0. In the problem at hand that factor is known to be negligible so that it can be concluded that the two methods give the same result in so far as the damping-time constant is concerned. Thus the previous variational calculations have been checked and the system transient response is given within the limitations of the methods.

Theoretical Analysis, Parallel Ferroresonant Case

STATIC-CIRCUIT CHARACTERISTIC

The circuit considered is shown in Fig. 2 where $R_L=R_0+R_{core}$.

$$i_L=\frac{\phi}{L_0}+\gamma\phi^3$$

$$i_c=C\frac{dv_1}{dT} \tag{35}$$

$$v_1=\frac{d\phi}{dT}+i_LR_L \tag{36}$$

$$v(T)=v_1+R_1i \tag{37}$$

Substituting equations 35 and 36 into equation 37 yields

$$\frac{d^2\phi}{dT^2}+\left(\frac{1}{R_1C}+\frac{R_L}{L_0}+3\gamma R_L\phi^2\right)\frac{d\phi}{dT}+\left(\frac{R_L}{R_1CL_0}+\frac{1}{L_0C}\right)\phi+\left(\frac{R_L\gamma}{R_1C}+\frac{\gamma}{C}\right)\phi^3=\frac{v(T)}{R_1C} \tag{38}$$

If the following substitutions are made, the differential equation 38 becomes almost identical to equation 4.

$$\frac{v(T)}{R_1C}=\frac{K}{R_1C}\cos\omega T-\frac{W}{R_1C}\sin\omega T=\frac{V_m}{R_1C}\cos(\omega T+\theta)$$

$$\alpha_1=-3\gamma R_L$$

$$\omega_{01}^2=\frac{R_L}{R_1CL_0}+\frac{1}{CL_0}$$

$$\beta_1=\frac{R_L\gamma}{R_1C}+\frac{\gamma}{C}$$

$$c_1=\frac{1}{R_1C}+\frac{R_L}{L_0}$$

Because of this similarity, it is not necessary to repeat the complete analysis again. Adapting equation 17 to the parallel case, the forcing voltage as a function of core flux is determined.

$$\frac{V_m^2}{(R_1C)^2}=A^6\left[\frac{4}{9}\left(\beta_1^2+\frac{\alpha_1^2\omega^2}{9}\right)\right]+A^4\left[\frac{4}{3}\beta_1(\omega^2-\omega_{01}^2)-\frac{c_1\alpha_1\omega^2}{3}\right]+A^2[(\omega^2-\omega_{01}^2)^2-(c_1\omega)^2] \tag{39}$$

where

$$\phi=A\cos\omega T-A/6(\cos 3\omega T)$$

Solving for total current $i=i_L+i_c$, from equations 35 and 36, gives

$$i=C\frac{d^2\phi}{dT^2}+\left[\frac{CR_L}{L_0}+3\gamma R_LC\phi^2\right]\frac{d\phi}{dT}+\frac{\phi}{L_0}+\gamma\phi^3 \tag{40}$$

Had a current source been used in Fig. 2, the solution would have been very similar to the series case with the voltage and current interchanging roles.

Solving for current, considering flux to be the independent variable, yields

$$\phi=A\cos\omega T$$

$$i=-\omega^2AC\cos\omega T+\left(\frac{CR_L}{L_0}+\frac{3\gamma R_LCA^2}{4}\right)\times(-\omega A\sin\omega T)+\frac{A}{L_0}\cos\omega T+\frac{3\gamma A^3}{4}\cos\omega T \tag{41}$$

$$i_{rms}^2=\left(\frac{\phi_{rms}}{L_0}+\frac{3}{2}\gamma\phi_{rms}^3-\omega^2C\phi_{rms}\right)+\left[\left(\frac{CR_L}{L_0}+\frac{3\gamma CR_L}{2}\phi_{rms}^2\right)\omega\phi_{rms}\right] \tag{42}$$

The negative-resistance characteristic can theoretically be traced out by applying a voltage source directly to the parallel ferroresonant circuit. This would force the flux in the core to be of some prescribed amplitude and because R_1 would then be zero, equation 39 would not apply. Thus a negative-resistance effect must be incorporated in equation 42, and the i versus ϕ plot must have a negative slope at some point if the jump phenomenon is to exist. To satisfy the negative-slope criterion $\omega L_0>1/\omega C$ and R_L must be small. This shows that the circuit must be capacitive near the origin. Using the same logic and reasoning here as in the series case, it can be concluded that the circuit passes through resonance at the "dip" point in the static circuit characteristic plot and becomes inductive with increasing voltage and current. The point of minimum current or maximum circuit impedance can then be used to calculate the parallel-circuit resistance at the "resonant" point.

As in the series case, the value of flux at the switching points can be calculated from equation 42. Allowing L_0 to be very large, the following results

$$\frac{d(i_{rms}^2)}{d(\phi_{rms}^2)}=\left(\frac{9}{4}\gamma^2+\frac{9}{4}\gamma^2C^2\omega^2R_L^2\right)3\phi_{rms}^4-6\gamma\omega^2C\phi_{rms}^2+\omega^4C^2=0 \tag{43}$$

$$\phi_{rms}^2=\frac{6\gamma\omega^2C\pm\sqrt{36\gamma^2\omega^4C^2-12\omega^4C^2\left(\frac{9}{4}\gamma^2+\frac{9}{4}\gamma^2C^2\omega^2R_L^2\right)}}{6\left(\frac{9}{4}\gamma^2+\frac{9}{4}\gamma^2C^2\omega^2R_L^2\right)} \tag{44}$$

The plus sign gives the flux at minimum current which is the resonant point.

$$\phi_{rms/min}=\sqrt{\frac{2\omega^2C}{3\gamma}},\ \text{for } R_L=0 \tag{45}$$

If $R_L\neq 0$, the value of calculated flux at

Table III. Series Resistance at Resonance

R_0 Value	Static-Circuit Characteristic, Ohms	External Measurements, Ohms
0	480	460
500	1,000	960

Table IV. Series Transition Speed

			$R_0=0$		$R_0=500$	
RMS, Volt-Second	i RMS, Milliamperes	Cycles Per Second	R, Ohms	τ, Milliseconds	R, Ohms	τ, Milliseconds
0.15	5	37	2,500	14	3,000	12
0.25	20	63	1,170	11	1,670	8
0.32	40	81	500	16	1,000	8
0.35	60	89	300	22	800	8
0.39	80	99	200	26		

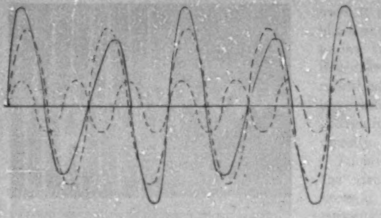


Fig. 9. The 90-cps damping frequency effect upon 60 cps

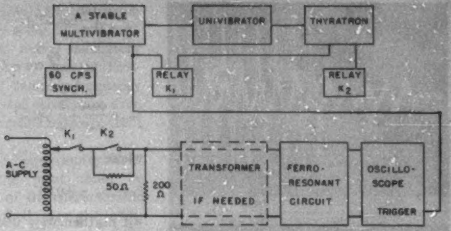


Fig. 11. Electronic circuitry for transition-speed observation

resonance will be slightly smaller than that indicated in equation 45.

The discriminate of equation 44 gives the maximum theoretical value of inductor series resistance.

$$R_{Lm} \approx \frac{1}{\sqrt{3}\omega C} \qquad (46)$$

As one would expect, equation 46 is identical with equation 11.

The third-harmonic correction can be determined by assuming $\phi = A\cos\omega T - A/6(\cos 3\omega T)$ in equation 40; see Appendix III. Assuming L_0 to be very large, the current corrected for third-harmonic flux becomes

$$i \approx \left(\frac{2}{3}\gamma A^3 - \omega^2 CA\right)\cos\omega T - \left(\frac{7}{12}\gamma R_L C\omega A^3\right)\sin\omega T \qquad (47)$$

Higher harmonics are neglected.

$$i_{rms}^2 = \left(\frac{16}{9}\gamma^2 + \frac{49}{36}\gamma^2 R_L^2 C^2 \omega^2\right)\varphi_{rms}^6 - \left(\frac{8}{3}\gamma\omega^2 C\right)\phi_{rms}^4 + (\omega^4 C^2)\phi_{rms}^2 \qquad (48)$$

The corrected maximum theoretical resistance is found to be the following

$$R_{Lm} \approx \frac{7}{8}\frac{1}{\sqrt{3}\omega C} \qquad (49)$$

This corrected value of R_{Lm} is within 3% of the experimental results. The correction of the third-harmonic flux component is noted to be rather large. Some reflection on this point reveals that because of the nonlinearities in the system, the circuit characteristic curve is different when one uses a voltage source from when one uses a current source. This is a very interesting phenomenon. When a voltage source is used as the forcing term, equation 39 does not apply because the flux is forced to be a single sine wave and R_1 equals zero. Equations 42 through 46 are valid and a very shallow negative-resistance region occurs. When a current source is used, it is equivalent to using a V_m and R_1 of large value. Note that the effect of R_1 on the right side of equation 39 is very small above some value of R_1 when $R_L/R_1/<<1$, assuming of course that V_m increases with R_1 to give the same current. Thus a current source allows harmonics of flux to occur and a deeper negative-resistance region results because the clipped flux peaks will cause less rms current to flow in the inductor for the same rms voltage.

This phenomenon is perhaps explained better by using graphical analysis. Referring to Figs. 3 and 4, the inductor saturation curve is somewhat different when obtained using a current source as when obtained by means of a voltage source. This difference causes the resonant points (*A*) and (*B*) to occur at different values of inductor current as well as different voltages. Because equivalent series resistance of the core is larger at lower currents, the curve in Fig. 4(B) has a lower over-all parallel-circuit resistance at resonance. Larger inductor currents make for less equivalent series inductor resistance or larger parallel resistance which will produce a deeper negative-resistance region for the curve in Fig. 4(A). This curve will also resonate at a slightly higher value of voltage. Fig. 4 is obtained by graphical addition

$$i = i_L + i_c$$

After the $R_L = 0$ curves have been constructed, a correction can be made for the values of R_L at different points, which is given in Fig. 4.

TRANSITION TIME

Because the form of the differential equation is the same for the parallel case as for the series case, the transient response is obtained by substituting the parallel-circuit values, α_1, β_1, C_1, and ω_{01} into the damping and oscillatory terms calculated previously.

$$\tau = \frac{2}{\frac{3\gamma R_L A^2}{2} + \frac{R_L}{L_0} + \frac{1}{R_1 C}} \qquad (50)$$

$$f_D = \frac{1}{4\pi} \times \sqrt{4\left[\frac{R_L}{R_1 C L_0} + \frac{1}{C L_0} + 3\left(\frac{R_L\gamma}{R_1 C} + \frac{\gamma}{C}\right)\frac{A^2}{2}\right] - \left[\frac{1}{R_1 C} + \frac{R_L}{L_0} + \frac{3\gamma R_L A^2}{2}\right]^2} \qquad (51)$$

Equations 50 and 51 apply to the flux during transition. In the parallel case

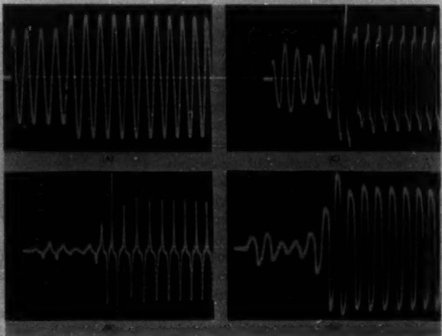


Fig. 10. Experimental series ferroresonant transition, C=1.0 μf, R₀=0

A—A-c supply voltage
B—Current transition
C—Voltage transition across inductor
D—Voltage transition across capacitor

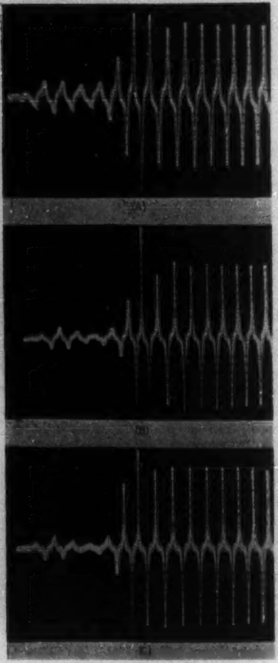


Fig. 12. Series ferroresonant transition; $R_0=0$. A-c supply same as Fig. 10

A—C = 0.5μf
B—C = 1.0μf
C—C = 2.0μf

the transition or discontinuity is a voltage phenomenon. Therefore, the flux response is identical with the voltage response except for a slight correction due to R_L. This correction is known to be small.

Experimental Verification

SERIES FERRORESONANT CASE

The accuracy of the theoretical analysis hinges on the accuracy of the assumed cubic saturation curve of the inductor; see Fig. 5. In obtaining this curve a voltage source was used and therefore the flux is forced to be sinusoidal.

$$\phi=\phi_p \cos \omega T=\frac{V_m}{\omega} \cos \omega T \tag{52}$$

From equation 3

$$I_p \cos \omega T=\frac{\phi_p}{L_0} \cos \omega T+\gamma\phi_p{}^3\times \left(\frac{3}{4}\cos \omega T+\frac{1}{4} \cos \omega T\right)\approx \frac{\phi_p}{L_0}\cos \omega T+\frac{3}{4}\gamma\phi^3 \cos \omega T \tag{53}$$

$$i_{rms}=\frac{\phi_{rms}}{L_0}+\frac{3}{2}\gamma\phi_{rms}{}^3 \tag{54}$$

where

$L_0=\infty$; $\gamma=0.85$

All mathematics used in this paper express $N\phi$ as ϕ so that no inconsistencies are encountered.

The equivalent series resistance of the core R_{core} was obtained by means of a special ultrasensitive wattmeter and conventional a-c milliammeter; see Fig. 6. The d-c winding resistance was 85 ohms. The reason for decreasing resistance with increasing current is explained by considering the size of the hysteresis loop compared with the square of the inductor current. Even though the power dissipated in the core increases with increasing current, that dissipated power does not increase as rapidly as $i_{rms}{}^2$, and the equivalent series resistance is a decreasing function of i. Theoretically, it should be possible to express this nonlinear resistance as a function of current and then proceed to solve a new nonlinear differential equation involving $f(i)$ for R_{core}. This matter has not been pursued and the present study deals with R in a quasi-linear fashion.

Another very simple method exists for approximating the over-all circuit characteristic. This graphical solution was mentioned previously in connection with the parallel case. For the series case it involves subtracting the inductive voltage from the capacitive voltage graphically, assuming zero resistance, and remembering that all rms quantities are positive. Then using load lines in the discontinuous region of different values, because of the nonlinear core resistance, the over-all circuit characteristic can be determined. Fig. 7 shows the composite results for the over-all circuit characteristic using the various techniques indicated. The reason for the rather large error in the graphically constructed curve is that harmonics are unaccounted for.

A check on the theoretical conclusions regarding the resonant point is obtained from Fig. 8 and is listed in Table III. This shows that the theoretical physical interpretation of the circuit operation is reasonably correct because if it were not the analyzed resonant point could not be truly resistive and therefore could not so accurately reflect the value of the external resistance R_0.

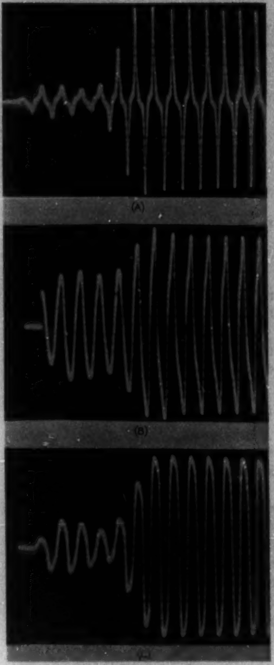


Fig. 13. Experimental series ferroresonant transition; $C=1.0\mu f$, $R_0=500$. A-c supply voltage waveform is identical with Fig. 10

A—Current transition
B—Voltage transition across reactor
C—Voltage transition across capacitor

TRANSITION TIME

The discontinuous transition time can be determined by a quasi-linear approach Calculation of the time constant and damping frequency at various values of flux allows one to view the transition in stages. Table IV gives τ and f_D according to the following equations obtained from equation 21.

$$\tau=\frac{2}{R\left(\frac{1}{L_0}+3\gamma\phi_{rms}{}^2\right)}=\frac{2}{3\gamma R\phi_{rms}{}^2} \tag{55}$$

$$f_D=\frac{\sqrt{3}\gamma\phi_{rms}}{4\pi}\sqrt{\frac{4}{C}-3\gamma\phi_{rms}{}^2R^2} \tag{56}$$

Applying a definition of response time as 3τ, after assuming an average value of τ, results in a transition speed of approximately $3^1/_2$ cycles for $R_0=0$, and a speed of about $1^1/_2$ cycles for $R_0=$

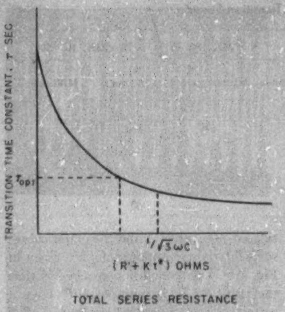


Fig. 14. Variation of transition time constant with lamination thickness

500.* Accounting for the damping frequency is very difficult because of its values exceeding the forcing frequency. The only thing that can be said is that if the time constant is small enough to damp out the discontinuous transition in approximately one cycle, no overshoot would be expected. It should be noted that the nonlinear effects in the time-constant evaluation tend to cancel themselves out when $R_0=500$. Another point is that the initial part of the transition for $R_0=0$ is rather fast, whereas, the end tends to slow down. Therefore, one would expect to observe the transition only slowly damping out near the end of its course. This effect is emphasized all the more by the oscillatory f_D contribution to the jump amplitude.

At large values of peak flux the value of f_D is in the neighborhood of 90 cps (cycles per second). Adding a 90-cycle wave to a 60-cycle wave, with the two in phase at the outset (see Fig. 9), and comparing this with the actual oscillatory response of Fig. 10 shows that the theoretical calculations of damping frequency have valid experimental correlation even though it seems impossible to relate them directly.

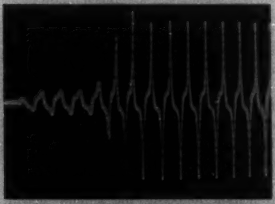


Fig. 15. Experimental series ferroresonant current transition with increased eddy-current loss. Secondary of reactor was loaded to simulate an increase in lamination thickness of approximately three times the actual thickness

To investigate the discontinuity itself the circuit operating point must be brought to the knee of the characteristic curve and then pushed over the edge. This was accomplished with the aid of the simple multivibrator-relay circuit shown in Fig. 11. The closing of K_2 was delayed a prescribed amount by adjustment of the univibrator. This delay must be sufficient to allow the ferroresonant circuit to settle down at the operating point just below the knee of the curve. A somewhat higher value of supply voltage is required as the external resistance R_0 is increased (C remaining constant). Changing the value of C, of course, changes the whole characteristic curve so that much higher supply voltages are required when C is raised and vice versa. The calculated values of transition damping-time constant are not affected directly by C but only indirectly by affecting the over-all circuit characteristic, which affects the value of R_{core}. However, these effects tend to cancel each other as can be seen in Fig. 12. Note that Figs. 10 and 13 show good agreement with the calculated results of Table IV. Actually, an external resistor of 47 ohms was used in all cases in order to view the current transition on the oscilloscope. Such a small value of resistance has negligible effect on the calculations.

Effect of Lamination Thickness

An approach to hysteresis and eddy-current effects is indicated by considering them represented as an equivalent series resistance. In general, it is known that this is not the best way to represent eddy-current losses; however, the equivalent series resistor representation lends itself directly to the equations at hand and is therefore advantageous.

According to the findings given in various references, the core losses due to eddy currents vary as the square of the laminations thickness.[7,8]

Consider R_{core} as divided into its hysteresis (R_h) and eddy-current (R_e) components. The current in the inductor, for the most part, is independent of the resistance but is determined by the value of inductance. Therefore, the eddy-current losses can be represented by a series resistance whose magnitude increases with increased losses.

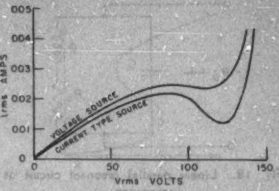


Fig. 16. Experimental static parallel-ferroresonant-circuit characteristics

$$R_e=Kt^2$$

where

R_e = eddy-current loss equivalent series resistance
K = constant, depending upon peak flux density and frequency
t = lamination thickness

Thus the total effective resistance in the circuit may be represented as

$$R=R_0+R_h+Kt^2=R'+Kt^2$$

Since the transition time varies inversely as the total effective resistance, the effect of lamination thickness on jump time can now be plotted; see Fig. 14. Previously, the maximum possible theoretical resistance was determined as being, $1/\sqrt{3}\omega C$; therefore, an upper limit on t^2 is indicated.

To check this result, a secondary-winding resistor was used to load down the inductor. This secondary loading has the same effect as allowing more eddy currents to flow, or increasing the thickness of laminations. Using the same wattmeter as previously and an audio oscillator, the core losses can be separated in the conventional manner. Because of the difficulty in reading fractions of a watt, this result is only approximate. However, at 60 cycles, the results indicate that the core losses are split about evenly. Therefore, increasing the secondary load until the watt loss doubled was approximately equivalent to increasing the eddy-current loss threefold or increasing the size of the laminations by $\sqrt{3}$. While this

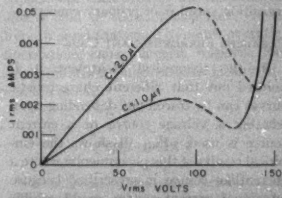


Fig. 17. Experimental static parallel-ferroresonant-circuit characteristic; $R_1=50K$

---

* The $\tau_{average}$ must be weighted toward the high-flux end because more time proportionately is spent in that region.

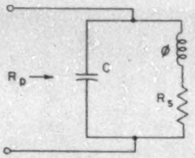


Fig. 18. Linear parallel resonant circuit at resonance

Table V. Parallel Transition Speed

		$R_1 = 100K$; $R_0 = 0$		$R_1 = 25K$; $R_0 = 0$	$R_1 = 25K$; $R_0 = 500$	
RMS, Volt-Second	Cycles Per Second	R_L, Ohms	τ, Milliseconds	τ, Milliseconds	R_L, Ohms	τ, Milliseconds
0.15	39	2,500	12	10	3,000	10
0.25	64	1,170	10	9	1,670	7
0.35	89	300	19	15	800	8
0.40	101	200	22	17	700	7

Note: $C = 1.0$.

is only approximate, at least a general order of magnitude of the effect of lamination thickness on transition speed can be determined. From equation 55, doubling the resistance R decreases the transition time by two. It is interesting to compare this eddy-current-effect case (see Fig. 15) with the situation when the external resistance R_0 is 500 ohms (Fig. 13). The calculations indicate a slightly slower transition speed for the increased core losses compared with adding an external resistance equal to the normal core loss at resonance.

It can be concluded that the effect of lamination thickness on transition time is somewhat critical. An increase in thickness will increase transition speed to the point where no transition at all occurs. As the magnitude of the resistance increases, the amplitude of the discontinuous part of the transition approaches zero. Therefore, there is a certain maximum value of lamination thickness in any particular case. Noting that the value of maximum theoretical resistance varies as $1/\omega C$, it follows that increasing C decreases the maximum possible series resistance. Thus, laminations must be thinner as C increases. For a particular case, τ_{opt} might be the optimum transition-time constant; see Fig. 14.

It should be pointed out that core losses are dissipated in the form of heat, and the danger exists of the inductor becoming too hot. If this is a problem, it would be better to dissipate the heat in an external resistor and design the inductor for minimum losses, assuming maximum possible transition speed is of primary concern.

PARALLEL FERRORESONANT CASE

In the theoretical treatment it is pointed out that different characteristic curves can be traced out depending upon whether a voltage source or a current source is used. Fig. 16 shows the empirical results of this phenomenon where a currentlike source is prescribed because an actual current source would exhibit the jump phenomenon. Therefore, the current source must not have an internal resistance greater than the value of negative resistance if the total characteristic curve is to be traced out. This unusual phenomenon was originally reported in a paper by Mackay without explanation and was discovered independently by the author.[9]

As in the series case, the dip in the characteristic curve indicates resonance, and the equivalent parallel resistance of the core can be determined at that point. It is noted that the equivalent resistance decreases when a larger capacitance is used; see Fig. 17. This is explained by the fact that, even though the series inductor resistance decreases because of higher current, the parallel resistance is affected by C. Considering the circuit linearized at resonance, the following expression can be applied (see Fig. 18)

$$R_p \approx \frac{1}{\omega^2 C^2 R_s}$$

Figs. 6 and 17 show that when the capacitance is doubled, the equivalent series resistance is approximately halved at resonance. Comparing these ideas with the empirical results, the indicated parallel resistance at resonance for $C = 1.0\mu f$ is 10,000 ohms as compared with the indicated parallel resistance for $C = 2.0\mu f$ which is 5,600 ohms. Thus the linear results at resonance carry over to the nonlinear case at resonance as one would expect.

TRANSITION SPEED

From equations 50 and 51

$$\tau = \frac{2}{3\gamma R_L \phi_{\text{rms}}^2 + \frac{1}{R_1 C}} \tag{57}$$

$$f_D = \frac{1}{4\pi} \times \sqrt{\frac{12\gamma\phi_{\text{rms}}^2}{C}\left(\frac{R_L}{R_1}+1\right) - \left(\frac{1}{R_1 C}+3\gamma R_L \phi_{\text{rms}}^2\right)^2} \tag{58}$$

where

$L_0 = \infty$, $\gamma = 0.85$
$C = 1.0\mu f$, $R_L = R_0 + R_{\text{core}}$

Three circuit configurations were considered, see Table V, and the empirical results for these computations are given in Fig. 19. One conclusion reached is that the value of R_1 is unimportant above a certain large value where the $1/R_1C$ terms become negligible. Another conclusion is that the value of R_L affects the transition time directly. Values of f_D are such that it is not possible to check directly their effect on jump time. However, at large values of flux, the damping frequency is approximately 90 cps, and from Figs. 9 and 19 it would appear that the calculations are reasonably correct. Obviously, if the time constant τ is small enough, no overshoot can result.

Eddy-current effects in the parallel case have the same influence as in the series case (Fig. 19). This is apparent from the "duality" of the systems as well as from the theoretical calculations which show that increased resistance in series with the inductor increases the transition speed until the amplitude of the discontinuity becomes zero.

SWITCHING SPEED IN BISTABLE SYSTEMS

Any device exhibiting bistable properties can be represented as having a negative-resistance characteristic; see Fig. 20. It is known that transitions in bistable systems can have varied transition speeds according to the pulse height of the independent variable x. Any particular system can be analyzed for this phenomenon, but there appears to be two general explanations which would apply to all cases. Detailed experimental verification of these apparently true concepts is proceeding.*

One explanation concerns the static fluctuations of the negative-resistance characteristic and the x pulse height. Both the characteristic and pulse height are undoubtedly undergoing small random or cyclic fluctuations. In order for a transition to occur, the x pulse height must be greater than the x coordinate of the negative-resistance characteristic knee (x_1, y_1). If a pulse height of x approximately equal to x_1 is applied, the transi-

---

* R. S. Mackay is presently engaged in this work.

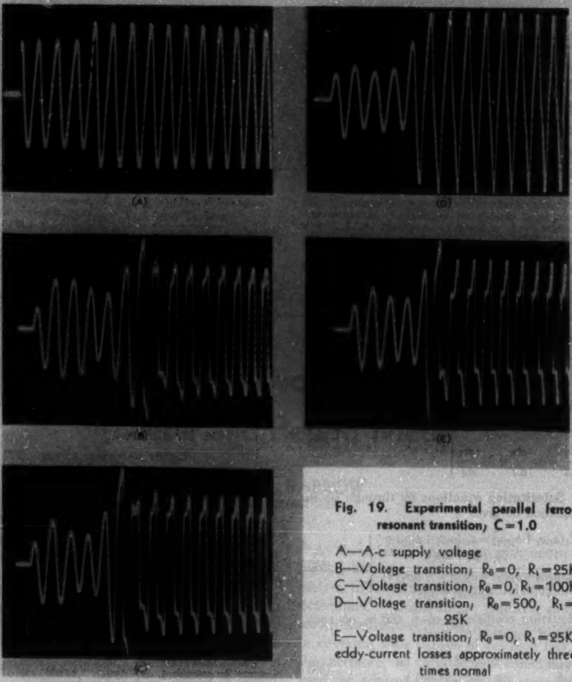


Fig. 19. Experimental parallel ferroresonant transition; C = 1.0

A—A-c supply voltage
B—Voltage transition; $R_0=0$, $R_1=25K$
C—Voltage transition; $R_0=0$, $R_1=100K$
D—Voltage transition; $R_0=500$, $R_1=25K$
E—Voltage transition; $R_0=0$, $R_1=25K$, eddy-current losses approximately three times normal

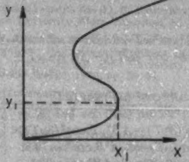


Fig. 20. General negative-resistance curve

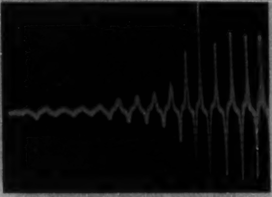


Fig. 21. Current transition for series-ferroresonant circuit with supply voltage of insufficient amplitude. A-c supply-voltage waveform is identical with Fig. 10(A)

tion in y might begin, but then a small fluctuation in either or both of the variables might tend to stop it. Thus the discontinuous jump would be slowed down until another fluctuation of opposite effect continued the process. When a slightly larger value of x is applied, only the larger fluctuations would affect the operation. Therefore, a faster transition speed would result. If the x pulse height is increased to a level above these minute fluctuations, a minimum transition time is achieved and increasing x any further does not result in a further decrease of transition time.

Another explanation with regard to the switching of stable states in bistable systems is the dynamic effect of an x pulse on the negative-resistance curve itself. In oscillatory systems, the effect of pulses can cause dynamic changes in the x-y characteristic which oscillate and damp out in time. These dynamic fluctuations of the negative-resistance curve under pulse conditions are different from the static fluctuations considered in the foregoing. The discontinuous transition speed would be affected the same way under either type of fluctuation. The transition would have a delay in getting started, and the speed would be considerably slower when pulse heights of x approximately equal to x_1 were applied as compared with x greater than x_1.

From experimental observations it would appear that the dynamic effects are the predominant ones in the ferroresonant cases. These dynamic effects appeared to be several times the magnitude of the static fluctuations.

In all the figures presented thus far the initiating voltage or current was sufficient to obtain maximum possible transition speed and thus determine the inherent limitations of the system. Fig. 21 does not show the slowest possible transition speed but only an intermediate value to indicate an easily attainable slow speed caused by insufficient voltage.

Conclusions

The static-circuit characteristic and transient response of simple ferroresonant circuits can be calculated directly from circuit parameters. To obtain good agreement between these results and empirical observations, it is necessary to include the nonlinear equivalent series core resistance in the analysis. There is a calculable maximum resistance in series with the inductor for which a discontinuity exists. The speed of transition is significantly affected by the nonlinear circuit resistance which was treated in a quasi-linear manner to give good results. The thickness of laminations of the inductor is somewhat critical; thicker laminations give faster transition speed (and smaller amplitude of discontinuity). Additional d-c resistance in series with the inductor also gives faster transition speeds in both the series and parallel cases while decreasing the amplitude of the discontinuity. The transition speed is relatively independent of the value of capacitance used because the capacitance effects tend to cancel themselves. It is hypothesized that small dynamic or static fluctuations cause sluggish response for small initiating pulse heights in any bistable system.

Appendix I

The solution of the series-ferroresonant differential equation including the third-harmonic component of flux is given in the following. Let $\phi=A\cos\omega T+B\cos 3\omega T$ in equation 4.

$$\phi^2\frac{d\phi}{dT}$$

$$\begin{aligned}&=(A^2\cos^2\omega T+2AB\cos\omega T\cos 3\omega T+\\&\quad B^2\cos^2 3\omega T)(-A\omega\sin\omega T-\\&\qquad 3\omega B\sin 3\omega T)\end{aligned}$$

$$\left.\begin{aligned}&=-\omega A^3\cos^2(\omega T)\sin(\omega T)-\\&\quad 2A^2B\omega\sin(\omega T)\cos(\omega T)\times\\&\quad\cos(3\omega T)-AB^2\omega T\cos^2(3\omega T)\times\\&\quad\sin(\omega T)-3\omega A^2B\cos^2(\omega T)\times\\&\quad\sin(3\omega T)-6\omega AB^2\cos\omega T\times\\&\quad\sin 3\omega T\cos 3\omega T-3\omega B^3\times\\&\qquad\cos^2 3\omega T\sin 3\omega T\end{aligned}\right\}\qquad(59)$$

$$\begin{aligned}\phi^3&=A^3\cos^3\omega T+3A^2B\cos^2\omega T\cos 3\omega T+\\&\quad 3AB^2\cos\omega T\cos^2 3\omega T+B^3\cos^3 3\omega T\end{aligned}\qquad(60)$$

The differential equation reduces to

$$\begin{aligned}&-\omega^2A\cos\omega T-9\omega^2B\cos 3\omega T-c\omega A\sin\omega T-\\&\quad 3\omega cB\sin 3\omega T+\omega_0^2A\cos\omega T+\\&\quad\omega_0^2B\cos 3\omega T\\&\quad=K\cos\omega T-W\sin\omega T+\alpha\phi^2\\&\qquad\qquad\frac{d\phi}{dT}+\beta\phi^3\\&\quad=K\cos\omega T-W\sin\omega T-\alpha A^3\omega\times\\&\quad\left(\frac{1}{4}\sin\omega T+\frac{1}{4}\sin 3\omega T\right)-\\&\quad\alpha\omega A^2B\left[\left(-\frac{1}{2}\sin\omega T+\frac{1}{2}\sin 5\omega T\right)-\right.\\&\quad\alpha\omega AB^2\left(\frac{1}{2}\sin\omega T-\frac{1}{4}\sin 5\omega T+\right.\\&\quad\left.\frac{1}{4}\sin 7\omega T\right)-3\alpha\omega A^2B\times\\&\quad\left(\frac{\sin\omega T}{4}+\frac{\sin 3\omega T}{2}+\frac{\sin 5\omega T}{4}\right)-\\&\quad 3\omega\alpha AB^2\left(\frac{\sin 5\omega T}{2}+\frac{\sin 7\omega T}{2}\right)-\\&\quad 3\alpha\omega B^3\left(\frac{\sin 3\omega T}{4}+\frac{1}{4}\sin 9\omega T\right)+\\&\quad\beta A^3\left(\frac{3}{4}\cos\omega T+\frac{\cos 3\omega T}{4}\right)+\\&\quad 3\beta A^2B\left(\frac{\cos\omega T}{4}+\frac{\cos 3\omega T}{2}+\right.\\&\quad\left.\frac{\cos 5\omega T}{4}\right)+3\beta AB^2\left(\frac{\cos\omega T}{2}+\right.\\&\quad\left.\frac{\cos 5\omega T}{4}+\frac{\cos 7\omega T}{4}\right)+\beta B^3\times\\&\qquad\left(\frac{3}{4}\cos 3\omega T+\frac{1}{4}\cos 9\omega T\right)\end{aligned}\qquad(61)$$

Setting $B=-A/6$ in equation 13. squaring and adding, gives

$$\begin{aligned}W^2+K^2&=\omega^2V_m^2=(\omega^2-\omega_0^2)^2A^2+\\&\quad\frac{4}{3}\beta(\omega^2-\omega_0^2)A^4+\frac{4}{9}\beta^2A^6+c^2\omega^2A^2-\\&\quad\frac{4}{9}c\alpha\omega^2A^4+\frac{4}{81}\alpha^2\omega^2A^6\\&=A^6\left[\frac{4}{9}\left(\beta^2-\frac{\alpha^2\omega^2}{9}\right)\right]+\\&\quad A^4\left\{\frac{4}{3}\left[\beta(\omega^2-\omega_0^2)-\frac{c\alpha\omega^2}{3}\right]\right\}+\\&\qquad A^2[(\omega^2-\omega_0^2)^2+c^2\omega^2]\end{aligned}\qquad(62)$$

Appendix II

Transformation of the perturbation differential equations to canonical form is as follows. Given the system

$$\left.\begin{aligned}\frac{dv_1}{dT}&=a_1v_1+b_1v_2\\\frac{dv_2}{dT}&=a_2v_1+b_2v_2\end{aligned}\right\}\qquad(63)$$

the canonical form is

$$\left.\begin{aligned}\xi&=\xi_0e^{S_1T}\\\eta&=\eta_0e^{S_2T}\end{aligned}\right\}\qquad(64)$$

Therefore, the linear transformation is

$$\left.\begin{aligned}\xi&=a_{11}v_1+b_{11}v_2\\\eta&=a_{22}v_1+b_{22}v_2\end{aligned}\right\}\qquad(65)$$

where

$$\begin{vmatrix}a_{11}&b_{11}\\a_{22}&b_{22}\end{vmatrix}\neq 0$$

$$\left.\begin{aligned}\frac{d\xi}{dT}&=a_{11}\frac{dv_1}{dT}+b_{11}\frac{dv_2}{dT}\\\frac{d\eta}{dT}&=a_{22}\frac{dv_1}{dT}+b_{22}\frac{dv_2}{dT}\end{aligned}\right\}\qquad(66)$$

Substituting equations 63 through 65 in equation 66 gives

$$\left.\begin{aligned}S_1(a_{11}v_1+b_{11}v_2)&=a_{11}(a_1v_1+b_1v_2)+\\&\quad b_{11}(a_2v_1+b_2v_2)\\S_2(a_{22}v_1+b_{22}v_2)&=a_{22}(a_1v_1+b_1v_2)+\\&\quad b_{22}(a_2v_1+b_2v_2)\end{aligned}\right\}\qquad(67)$$

Equating coefficients of v_1 and v_2 on each side of equation 67 gives

$$\left.\begin{aligned}a_{11}(a_1-S_1)+b_{11}a_2&=0\\a_{11}b_1\qquad+b_{11}(b_2-S_1)&=0\end{aligned}\right\}\qquad(68)$$

$$\left.\begin{aligned}a_{22}(a_1-S_2)+b_{22}a_2&=0\\a_{22}b_1\qquad+b_{22}(b_2-S_2)&=0\end{aligned}\right\}\qquad(69)$$

Solving for a_{11}, b_{11}, a_{22}, and b_{22}, it is seen that if any results are to be attained, S_1 and S_2 must satisfy the following determinate which gives the characteristic equation:

$$\begin{vmatrix}a_1-S&a_2\\b_1&b_2-S\end{vmatrix}=0\qquad(70)$$

Appendix III

The solution of i in the parallel case including the third harmonic of flux is as follows. Let $\phi=A\cos\omega T-A/6(\cos 3\omega T)$ in equation 40.

$$\begin{aligned}i&=-\omega^2CA\cos\omega T+\frac{3}{2}AC\omega^2\cos 3\omega T-\\&\quad\frac{CR_L\omega A}{L_0}\sin\omega T+\frac{CR_L\omega A}{2L_0}\sin 3\omega T+\\&\quad 3\gamma R_LCA^3\omega\left(\cos^2\omega T\sin\omega T+\right.\\&\quad\frac{\cos^2\omega T\sin 3\omega T}{2}+\\&\quad\frac{\sin\omega T\cos\omega T\cos 3\omega T}{3}+\\&\quad\frac{\cos\omega T\cos 3\omega T\sin 3\omega T}{6}+\\&\quad\frac{\cos^2 3\omega T\sin\omega T}{36}+\\&\quad\left.\frac{\cos^2 3\omega T\sin 3\omega T}{72}\right)+\frac{A}{L_0}\cos\omega T-\\&\quad\frac{A}{6L_0}\cos 3\omega T+\gamma A^3\left(\cos^3\omega T-\right.\\&\quad\frac{\cos^2\omega T\cos 3\omega T}{2}+\\&\qquad\left.\frac{\cos\omega T\cos^2 3\omega T}{12}+\frac{\cos^3 3\omega T}{216}\right)\end{aligned}\qquad(71)$$

All of the trigonometric terms can now be evaluated as a sum of single-harmonic terms with the aid of a trigonometric identity table.

$$\begin{aligned}i&=\left(-\omega^2AC+\frac{A}{L_0}+\frac{3}{4}\gamma A^3-\frac{\gamma A^3}{8}+\frac{\gamma A^3}{24}\right)\times\\&\quad\cos\omega T+\left[-\frac{\omega ACR_L}{L_0}+3\gamma R_LCA^3\omega\times\right.\\&\quad\left.\left(-\frac{1}{4}+\frac{1}{8}-\frac{1}{12}+\frac{1}{72}\right)\right]+\\&\qquad\text{higher harmonic terms}\end{aligned}\qquad(72)$$

References

1. Nonlinear Vibrations in Mechanical and Electrical Systems (book), J. J. Stoker. Interscience Publishers Inc., New York, N. Y., 1950.

2. On the Synthesis of Nonlinear Circuits, L. M. Vallese. *Proceedings*, Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, Brooklyn, N. Y., vol. II, 1953, pp. 201–14.

3. Ferro Inductance as a Variable Electric-Circuit Element, J. D. Ryder. *AIEE Transactions*, vol. 64, Oct. 1945, pp. 671–78.

4. Introduction to Non-Linear Mechanics (book), N. Minorsky. Edwards Bros. Inc., Ann Arbor, Mich., 1947, pp. 42–45; 186–87.

5. Forced Oscillations in Nonlinear Systems (book), C. Hayashi. Nippon Printing and Publishing Company, Ltd., Tokyo, Japan, 1953, pp. 81–85.

6. Introduction to Nonlinear Mechanics (book), N. Kryloff, N. Bogoliuboff (translation by S. Lewschitz), Princeton University Press, Princeton, N. J., 1943.

7. Handbook of Engineering Fundamentals (book), O. W. Eshback. John Wiley & Sons, Inc., New York, N. Y., 1952, pp. 9–33.

8. Soft Magnetic Materials Used in Industry (book), A. E. DeBarr. Institute of Physics, London, England, 1953, p. 18.

9. Sinusoidal and Relaxation Oscillations Sustained by Nonlinear Reactances, R. S. Mackay. *Journal of Applied Physics*, New York, N. Y., vol. 24, no. 9, Sept., 1953, pp. 1163–67.

10. Interesting Nonlinear Effects, R. S. Mackay. *Ibid.*, no. 3, Mar. 1953, pp. 311–13.

11. An Approximate Graphical Analysis of the Steady-State Response of Nonlinear Networks, S. Duinker. *Philips Research Report*, Amsterdam, Netherlands, no. 8, Apr. 1953, pp. 133–44.

12. A Shifting Register Using Ferroresonant Flip Flops, S. Duinker. *Applied Scientific Research*, The Hague, Netherlands, vol. 4B, 1955, pp. 317–28.

13. Ferroresonant Flip Flop, C. Isborn. *Electronics*, New York, N. Y., vol. 25, no. 4, Apr. 1952, pp. 121–23.

14. Switching in Bistable Circuits, R. S. Mackay. *Journal of Applied Physics*, New York, N. Y., vol. 25, no. 4, Apr. 1954, pp. 424–29.

15. Oscillatory Circuits Containing Iron-Cored Inductances, Abd El-Samie Mostafa, M. El-Shishini. *AIEE Transactions*, vol. 72, pt. I, July 1953, pp. 274–83.

16. Forced Oscillations with Nonlinear Restoring Force, C. Hayashi. *Journal of Applied Physics*, New York, N. Y., vol. 24, no. 2, Feb. 1953, pp. 198–207.

17. Stability Investigation of the Nonlinear Periodic Oscillations, C. Hayashi. *Ibid.*, no. 3, Mar. 1953, pp. 344–48.

18. Acceleration Plane Method for Analysis of a Circuit with Nonlinear Inductance and Nonlinear Capacitance, Y. H. Ku. *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 619–26.

19. A Mathematical Analysis of a Series Circuit Containing a Nonlinear Capacitor, Louis A. Pipes. *Ibid.*, vol. 72, pt. I, July 1953, pp. 238–44.

20. Operational Methods in Nonlinear Mechanics, L. A. Pipes. *Report 51–10*, Department of Engineering, University of California at Los Angeles, Los Angeles, Calif., 1951.

21. On Extraordinary Phenomena in a Nonlinear Forced Oscillation, A. Sakurai. *Journal of the Physical Society of Japan*, Tokyo, Japan, vol. 10, Apr. 1955, pp. 274–78.

22. A Graphical Sinusoidal Analysis of a Nonlinear RC Phase Shift Feedback Circuit, D. Barbiere. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 43, no. 6, June 1955, pp. 679–84.

23. General Method of Solving Second Order Differential Equations by Phase Displacements, L. S. Jacobsen. *Journal of Applied Mechanics*, New York, N. Y., vol. 19, Dec. 1952, p. 543.

24. Contributions to the Theory of Nonlinear Oscillations (book), S. Lewschitz, editor. Princeton University Press, Princeton, N. J., 1950.

25. On a Forced Discontinuous Oscillation, Z. Nagumo. *Proceedings*, Fujihara Memorial Faculty of Engineering, Keio University, Tokyo, Japan, vol. 7, no. 25, 1954, pp. 36–43.

26. On Intermittent Oscillation and Oscillation Hysteresis, Z. Nagumo. *Ibid.*, no. 27, 1954, pp. 72–82.

27. On Mackay's Nonlinear Circuit, Z. Nagumo. *Ibid.*, vol. 6, no. 20, 1953, pp. 1–9.

28. Graphical Analysis of Nonlinear Circuits Using Impedance Concepts, J. S. Thomsen. *Journal of Applied Physics*, New York, N. Y., vol. 24, no. 11, Nov. 1953, p. 1379.

29. An Analysis and Design of the Ferroelectric Resonant Trigger Pair, C. E. Gremer. *Proceedings*, National Electronics Conference, Chicago, Ill., 1955, pp. 828–37.

The Logarithmic-Diode Counting-Rate Meter and Period Meter

BRUCE B. BARROW
NONMEMBER AIEE

Synopsis: This paper analyzes the logarithmic-diode counting-rate meter and the logarithmic-diode period meter, treating the problem of obtaining rapid response to reactor-flux transients while minimizing the spurious fluctuations caused by the random character of neutron detections in a fission counter. A procedure for choosing design parameters is discussed.

AUTOMATIC start-up of nuclear reactors becomes increasingly important as reactors are designed for use in ships and other mobile power plants. Many applications, especially military ones, demand the fastest start-up that is safe for the reactor and for the operating personnel. To control rapid start-up requires that transients in neutron flux be measured at the lowest levels possible, levels corresponding to 10^{-10}, or less, of full operating power.

At these levels one works with instruments, like counters and counting-rate meters, that operate on individual neutron detections. Logarithmic counting-rate meters are frequently used for convenience in covering wide ranges of neutron flux. Furthermore, the variable most commonly used for control during reactor start-up is the time derivative of the logarithm of the neutron flux,[1] which is proportional to the reciprocal of the period (defined in equation 1). It is thus a direct step to attempt to measure period at low flux levels by differentiating the output of a logarithmic counting-rate meter, especially since period has conventionally been measured at higher flux levels by differentiating the logarithm of the output of an ion chamber.[2,3]

The special problem that arises when a period meter is built around a logarithmic counting-rate meter is random fluctuation of the meter output, or noise. This noise is inseparably associated with the quantity being measured, for it arises in the randomness with which neutrons are detected in the fission counter.

There is no possibility therefore of completely eliminating this noise. More accurate determination of period requires examination of a larger number of neutron detections, which in turn requires measurement over a longer interval of time, assuming that sensitivity of the fission counter can not be increased. The noise can, however, be minimized, and the general problem of obtaining suitably short response time with suitably accurate measurement is the central theme of this paper.

The following sections treat, first, the noise in a logarithmic-diode period meter; second, the linear circuitry associated with such a meter; third, the logarithmic-diode counting-rate meter, which comprises the "front end" of a logarithmic-diode period meter; and last, the choice of design parameters for a logarithmic-diode period meter.

Noise in Logarithmic-Diode Period Meter

General Operation

Fig. 1 represents, in block diagram, the important elements of the logarithmic-diode period meter. The system operates in the following manner: A small fraction of the neutrons that pass through the volume occupied by the fission counter produce current pulses at the fission-counter output. The customary assumption is made that the counting rate is proportional to the number of neutrons transiting the detector per unit time and is therefore a direct measure of the neutron flux within the reactor at the detector location.* It is assumed that pulses due to background alpha particles are rejected by the pulse-height discriminator. Each pulse in the discriminator output therefore represents a neutron detection and causes the pulse-forming circuit to inject a charge into the logarithmic smoothing circuit.

The logarithmic filter smooths the pulses into a continuous current, which varies randomly about a value propor-

---

Paper 57-403, recommended by the AIEE Nucleonic and Radiation Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE 2nd Nuclear Engineering and Science Conference, Philadelphia, Pa., March 11–14, 1957. Manuscript submitted December 5, 1956; made available for printing June 19, 1957.

Bruce B. Barrow is with Hycon Eastern, Inc., Cambridge, Mass.

Nearly all of the work presented in this paper was reported in a thesis of the same title submitted in 1955 to the Massachusetts Institute of Technology in partial fulfillment of the requirements for a graduate degree in electrical engineering. The author would like to express appreciation to Prof. T. S. Gray, who supervised the thesis, and to William M. Grim, Jr., and J. C. Simons, Jr., all of whom contributed valued advice. The work was made possible through support extended the M.I.T. Servomechanisms Laboratory by the Office of Naval Research and the Navy Bureau of Ships.

* The term "counting rate" as here used refers to a parameter of the idealized, mathematical model of the reactor and fission-counter system. The counting rate of a physical system can never be measured with perfect accuracy in any observation of finite length. The output of any counting-rate meter is therefore only a statistical estimate of the quantity that is here called the counting rate.

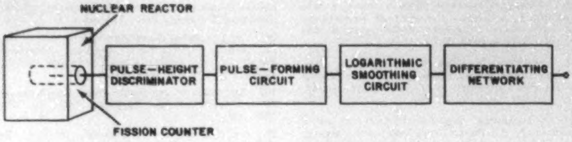


Fig. 1. Block diagram of period-measuring system for use at low flux levels

tional to the counting rate, and generates a voltage proportional to the logarithm of this current. This voltage is electrically differentiated, and the result is displayed on a suitably calibrated meter or, in an automatically controlled reactor, used as a control signal in a feedback loop. Practical performance falls short of the ideal not only because of the random character of the raw data, but also because of the time lags in the smoothing and differentiating circuits.

To discuss the statistical properties of the fluctuating period-meter output requires first a definition of the output quantity to be measured. When neutron flux is constant, for example, the true period of the reactor is infinite, and the measured period fluctuates randomly about infinity. For this reason it becomes almost meaningless to consider the statistics of the measured period. The voltage (or current) output of the differentiating circuit, however, is proportional to the reciprocal of the measured period. Because this output voltage is one signal used in automatic control of reactors, since it is statistically well behaved, and a variable proportional to the period reciprocal has been defined and given the name "growth factor." It will be remembered that the period T is defined[4] by the equation

$$T^{-1}=d(\ln P)/dt=\dot{P}/P \tag{1}$$

where P is some variable proportional to neutron flux. (Here and in what follows, a dot above a variable denotes differentiation with respect to time.) The growth factor g is defined by the equation

$$g(t)=d(\log_b P)/dt=(\log_b e)\dot{P}/P \\ =(\log_b e)/T \tag{2}$$

The quantity b, the base of logarithms, provides a scale factor and may be chosen as convenient. If b is the Naperian base, g is equal to the period reciprocal, and would be expressed in nepers per unit time. A more convenient base for engineering work is $10^{0.1}$, which gives

$$g(t)=d(\log_{10^{0.1}} P)/dt=10\, d(\log_{10} P)/dt \\ =4.3429\ \dot{P}/P \text{ decilogs per unit time} \tag{3}$$

The decilog[5] is a logarithmic measure of the same size as the decibel defined to describe power ratios. Its convenience comes from the fact that 10 decilogs of growth is equivalent to multiplication by a factor of 10, and 3 decilogs, very nearly equivalent to a factor of 2. The time required for 3 decilogs of growth is thus the doubling time.

EQUATIONS OF OPERATION

Fig. 2 shows the period-computing circuits in a simple meter. The input current i to the logarithmic smoothing circuit consists of very short current pulses each carrying charge Q, and n of these pulses occur on the average per second, producing an average current nQ. This average current, as noted before, is considered proportional to the reactor flux.

The diode in Fig. 2(B) operates in its logarithmic region with a current-voltage characteristic given by the equation

$$e_b=A \ln Bi_b \tag{4}$$

where A and B are constants. The capacitance C is made large enough so that the variations in plate current caused by the random character of the neutron detections are small compared to the magnitude of the plate current itself. For small perturbations in plate current the diode may be replaced by a linear resistor equal to the incremental resistance of the diode at its point of operation. This incremental resistance is given by

$$R_1=de_b/di_b=A/i_b=A/nQ \tag{5}$$

The output of the logarithmic smoothing circuit is differentiated electrically and, in this system, displayed on an indicating meter. The indicating meter here performs the useful auxiliary function of smoothing the output of the differentiator. It will be shown later that this function is so important that if, as in a feedback control loop, no indicating meter is provided, the general differentiating network shown in Fig. 1 should be designed to attenuate the higher frequencies.

All three networks that make up the period-computing circuits are linear for small increments, and may be described for small perturbations by stating the differential equations that relate input and output. If the three are isolated and system gains are neglected, then

$$R_1C\dot{e}_1+e_1=R_1i \tag{6}$$

$$T_2\dot{e}_2+e_2=T_2\dot{e}_1 \tag{7}$$

$$T_3\dot{\gamma}+\gamma=e_2/AT_2 \tag{8}$$

where T_2 stands for R_2C_2, and i and e_1 represent increments in the corresponding total quantities. Equation 8 is assumed to describe the time behavior of a simple indicating meter; the factor AT_2 in this equation is required to calibrate the meter output in terms of γ, the indicated growth factor.

NOISE IN THE LOGARITHMIC-DIODE PERIOD METER

The pulses generated by the pulse-forming circuit comprise a type of random process known as shot noise, because the distinguishing feature of shot noise is that it is composed of the superposition of many individually independent pulses. Shot noise has been extensively studied[6,7] and to the extent that the period-computing circuits operate linearly for small increments, the results of shot-noise theory may be applied to study the random fluctuations in e_1, e_2, and γ. The method is applicable, in particular, if the model defined by equations 6, 7, and 8 is realistic. Clearly neither the shot-noise

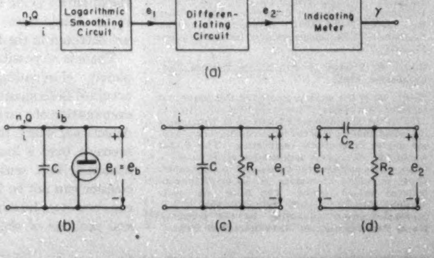


Fig. 2. Period-computing circuits

A—Block diagram
B—Simplified logarithmic smoothing circuit
C—Linear approximation to logarithmic smoothing circuit
D—Simple differentiating circuit

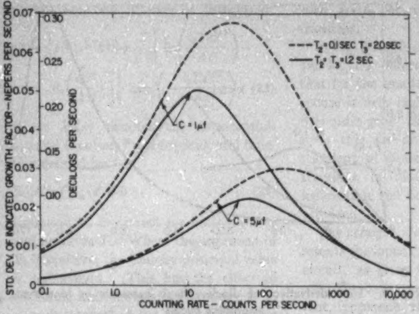


Fig. 3. Calculated curves illustrating the standard deviation of indicated growth factor as a function of counting rate

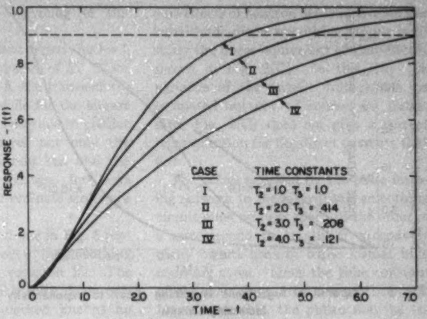


Fig. 4. Transient response of various differentiating networks to unit ramp input

model nor the linear circuit model can be used to calculate the d-c value of e_1, for instance, for that value depends upon the nonlinear properties of the diode.

A study of the random variations in γ has been carried out for the special case in which neutron flux is constant (i.e., for infinite period or zero growth factor), and the results, both experimental and theoretical, have been published.[8] For this special case, the mean value of γ is zero, and the variance (mean square deviation about the mean) is given by

$$\operatorname{var}(\gamma)=nQ^2/2(AC+nQT_2)(AC+nQT_3)(T_2+T_3) \quad (9)$$

in nepers² per sec² (second). Setting T_2 equal to T_3 gives a special case of practical interest

$$\operatorname{var}(\gamma)=nQ^2/4T_2(AC+nQT_2)^2 \quad (10)$$

Any randomness in Q, the charge in a current pulse, will increase the statistical errors of measurement. If the pulse-forming network is eliminated, or if for any other reason randomly varying charges are dumped into the logarithmic smoothing circuit, the generalized equation that applies can be obtained by replacing Q^2 and Q in equation 9 by the mean square and mean values of charge respectively.

Curves derived from equation 9 are shown in Fig. 3. These curves are based on a meter for which

$Q=5.9\times10^{-9}$ coulomb

$A=0.1$ volt per neper

The experimental results referred to in the foregoing were made with a meter having $T_2=0.1$ and $T_3=2.0$, and they agree with the dashed-line curves of Fig. 3.

The analysis given later shows that the results just presented can be applied, in practice, almost directly to the case with neutron flux transients, i.e., with finite period or nonzero growth factor.

Response of Differentiating Network

PROPERTIES REQUIRED

The differentiating network must perform a dual function: it must differentiate the low-frequency portion of the voltage output of the logarithmic diode to provide growth-factor information, and it must reject the high-frequency fluctuations or noise.

The simplest network that performs both these functions is provided by the combination of simple differentiating circuit and indicating meter discussed in the preceding section. The simple differentiating circuit is essentially a high-pass filter, while the indicating meter is a low-pass filter. Together they form a band-pass filter that differentiates frequencies lying well below its lower cutoff frequency and attenuates frequencies lying well above its upper cutoff frequency. Only that band of frequencies that is differentiated can supply the growth-factor information. If a band of frequencies is not differentiated it contributes only noise, and should therefore be rejected. From this argument it is clear that the two cutoff frequencies should not be far apart. The following analysis shows that for this type of differentiating circuit to provide the best noise filtering consistent with a particular speed of response, the two time constants T_2 and T_3 should be equal.

PERFORMANCE AS NOISE FILTER

It will be shown later that the frequency spectrum of e_1 becomes wider and wider with increased counting rate. For the differentiating network to have any effect whatever as a noise filter, this spectrum must extend above its upper cutoff frequency, at least at high counting rates. The efficacy of a differentiating network as a noise filter may therefore for this application be measured by calculating its response to white noise, i.e., wide-band noise with a spectrum that is flat over all frequencies of interest.

The fundamental relationship between the input and output power density spectrums of a linear network is[9]

$$\Phi_{oo}(\omega)=\Phi_{ii}(\omega)\left|H(j\omega)\right|^2 \quad (11)$$

where $\Phi_{oo}(\omega)$ is the output spectrum; $\Phi_{ii}(\omega)$ the input spectrum; and $H(j\omega)$ the complex frequency-response function of the network. For white-noise input, the input spectrum may be taken to be unity. Then, for the frequency response function defined by equations 7 and 8, (neglecting gain constants, which are irrelevant here)

$$\begin{aligned}\Phi_{oo}(\omega)&=\left|j\omega/(1+j\omega T_2)(1+j\omega T_3)\right|^2\\&=\omega^2/(1+\omega^2T_2^2)(1+\omega^2T_3^2) \quad (12)\end{aligned}$$

The mean square value of the noise that passes through the filter is given by integrating equation 12 over all frequencies:

$$\int_0^\infty \Phi_{oo}(\omega)d\omega=\pi/2T_2T_3(T_2+T_3) \quad (13)$$

Equation 13 may be interpreted to mean that all pairs of T_2 and T_3 that satisfy the equation

$$T_2T_3(T_2+T_3)=c \quad (14)$$

for some particular value of c, represent differentiating networks (of this particular type) that are equally effective as noise filters.

TRANSIENT RESPONSE

When the logarithmic smoothing circuit is properly indicating a constant growth

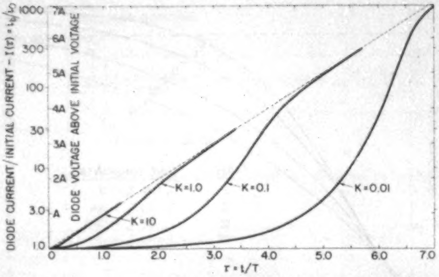


Fig. 5. Response of logarithmic smoothing circuit to exponentially increasing currents

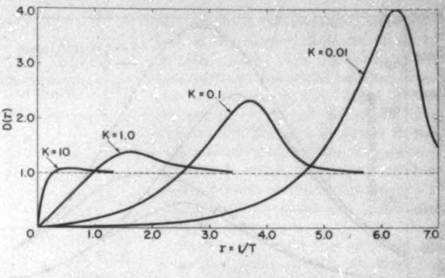


Fig. 6. Time derivatives of diode-voltage transients in logarithmic smoothing circuit

factor, its output voltage will be linearly increasing, except for the random fluctuations. For this application the transient response of a differentiating network may therefore best be evaluated by considering its response to a ramp input

$$r(t)=0,\ t\leq 0$$
$$r(t)=t,\ t\geq 0 \qquad (15)$$

With this input applied to a system defined by equations 7 and 8, and again neglecting gain constants, the output is

$$f(t)=1-(T_2\epsilon^{-t/T_2}-T_3\epsilon^{-t/T_3})/(T_2-T_3) \qquad (16)$$

when $T_2\neq T_3$, and

$$f(t)=1-(1+t/T_2)\epsilon^{-t/T_2} \qquad (17)$$

when $T_2=T_3$.

Fig. 4 shows $f(t)$ plotted for various pairs of time constants that satisfy equation 14 for $c=2$. If it is remembered that all four of the cases shown are equally effective as noise filters, then it is clear that case 1 must be chosen over the others for its superior transient response. Choosing a different value for the constant c in equation 14 would have no effect upon this analysis except to change the time scale. It follows that, whatever amount of noise filtering is desired, a well-designed period meter of the simple form discussed earlier will have $T_2=T_3$.

If the response time is arbitrarily defined to be the time required for $f(t)$ to reach 0.9, then differentiating networks of equally good response time may be compared for their effectiveness as noise filters. By this definition the two networks used for the calculations of Fig. 3, with

$T_3=0.1$ sec, $T_2=2.0$ sec

and $T_2=T_3=1.2$ sec

both have a response time of 4.7 sec, and are thus equally fast as differentiators of input ramps. Fig. 3 shows their vastly different performance as noise filters, and again the clear superiority of the differentiating network with equal time constants is demonstrated.

SUMMARY

A differentiating network that consists only of the simple circuit of Fig. 2(D) will not provide reduced period-meter noise at the high counting rates. This may be shown by setting T_3 equal to zero in equation 9.

A differentiating network of the form of the simple band-pass filter analyzed in the two preceding sections should have both time constants equal.

A differentiating network with sharper cutoff at the high frequencies could be designed to show some improvement over the solid-line curves of Fig. 3, while maintaining an equally good response time.

When the form of the differentiating network has been decided, and the optimum internal relationship among time constants and damping coefficients determined, the period-meter designer must balance filtering ability against response time to satisfy his particular requirements.

Analysis of Logarithmic-Diode Smoothing Circuit

The logarithmic-diode smoothing circuit, shown in Fig. 2(B), has a number of interesting properties that make it useful both as a wide-range counting-rate meter and as a front end for a low-level period meter. Nevertheless, its operation does not seem generally to be understood, and it has not heretofore received wide attention in the published literature. The following analysis presents first a general solution to the circuit equation, then specific solutions showing the transient response of the circuit to increasing exponentials, and finally certain results relating to the noise produced by randomly occurring neutron detections.

GENERAL SOLUTION OF THE CIRCUIT EQUATION

From Kirchhoff's current law, the equation describing the network of Fig. 2(B) is

$$C\dot{e}_1+i_b=i \qquad (18)$$

This gives, when the logarithmic-diode relationship of equation 4 is used, the following equation

$$CA\ d(\ln Bi_b)/dt+i_b=CA\dot{i}_b/i_b+i_b=i \qquad (19)$$

Equation 19, expressing diode current in terms of input current, is a nonlinear differential equation with no obvious solution. If, however, the substitution $z=1/i_b$ is made, equation 19 becomes

$$CA\dot{z}+iz=1 \qquad (20)$$

Equation 20 is not linear in i and z together, but if i is a specified function of time, it may be regarded as a linear differential equation (in z) with time-varying coefficient, and a solution valid for the particular driving current is direct. Because the equation is not linear in i and z together, however, superposing two different i's does not produce a solution that is the superposition of the solutions to the separate inputs.

The general solution to equation 20 for a specified input $i(t)$ is

$$z(t)=\left\{z(0)+1/AC\int_0^t \exp[(1/AC)\times \int_0^w i(y)dy]dw\right\}\exp[(-1/AC)\int_0^t i(w)dw] \qquad (21)$$

where w and y are variables of integration.

RESPONSE TO EXPONENTIAL INPUTS

If, at $t=0$, a transient of the form

$$i(t)=i_0\epsilon^{t/T} \qquad (22)$$

is applied to the logarithmic smoothing

circuit, equation 21 may be integrated:

$$z(t)=\left\{z(0)\epsilon^{i_0T/AC}+\frac{T}{AC}\left[\mathrm{E}_1\left(\frac{i_0T\epsilon^{t/T}}{AC}\right)-\mathrm{E}_1\left(\frac{i_0T}{AC}\right)\right]\right\}\exp\left(-\frac{i_0T\epsilon^{t/T}}{AC}\right) \qquad (23)$$

where $\mathrm{E}_1(x)$ denotes the exponential-integral function,[10] a transcendental function defined by

$$\mathrm{E}_1(x)=\int_{-\infty}^{x}(\epsilon^t/t)dt \qquad (24)$$

Equation 23 is correct for both positive and negative T. Where the argument of E_1 is positive, the Cauchy principal value is understood. (This has the effect of excluding from integration in equation 24 a small interval of the t-axis centered on the origin.) The exponential-integral function is well tabulated.[11]

For a systematic study of the results implied in equation 23, it is convenient to introduce several dimensionless parameters. Two of these are obvious:

$$\tau\equiv t/T \qquad (25)$$

$$I\equiv i_b/i_0=1/i_0z \qquad (26)$$

The first is merely time expressed as the number of periods that have passed. The second is a dimensionless response ratio, diode current at time t divided by initial input current. The third parameter is more complex, for it lumps together four of the fundamental quantities of the equation. This parameter

$$K\equiv i_0T/AC \qquad (27)$$

expresses, in a way, the adequacy of the system to respond to transients of a given period. The larger the constant K, the better the response of the logarithmic smoothing circuit. A large T implies a slow transient, and a large i_0 implies a large input current to force rapid transient response. On the other hand, a large capacitor C diverts much of the input current from the diode, and a large diode constant A means that a small current change in the diode produces a large voltage change and thus a large change in the charge stored on C.

For the general case a fourth dimensionless parameter is required, the ratio of i_0, initial input current, to $i_b(0)$, initial diode current:

$$\lambda\equiv i_0/i_b(0)=i_0z(0) \qquad (28)$$

Now equation 23 can be rewritten

$$1/I(\tau)=\epsilon^{-K\epsilon^\tau}[\lambda\epsilon^K-K\mathrm{E}_1(K)+K\mathrm{E}_1(K\epsilon^\tau)] \qquad (29)$$

Various cases with $\lambda=1$ offer considerable insight into the behavior of the circuit. These are the cases in which the logarithmic smoothing circuit is in equilibrium, with $i_0=i_b$, at the beginning of the transient.

In Fig. 5 $I(\tau)$ is plotted versus τ for $\lambda=1$ and for four different values of K. Note that for the smallest K the transient response is very bad, while for the largest it is quite good. It is possible to predict from this set of curves not only the transient in diode current but also the transient in diode voltage, for both logarithmic and linear ordinate scales are provided.

The straight, dashed line in Fig. 5 represents the input current to the smoothing circuit, as given by equation 22. The solid-line curves, which show diode current, approach the dashed line as an asymptote (i.e., the logarithm of diode current approaches the logarithm of input current). The difference between input current and diode current is, of course, the capacitor current, which is proportional to the derivative of diode voltage.

It is useful to study this derivative to determine which limitations of the period meter inhere in the logarithmic smoothing circuit. In the absence of noise an extremely rapid differentiating circuit could be built, one that would provide "instantaneous" derivatives of diode voltage. Fig. 6, in which is plotted $D(\tau)$, the logarithmic derivate of $I(\tau)$, indicates the shortcomings of the logarithmic smoothing circuit.

$$D(\tau)\equiv\frac{d}{d\tau}\ln I(\tau)=\frac{1}{I(\tau)}\frac{dI}{d\tau} \qquad (30)$$

This set of curves emphasizes the time delays, which were obvious in Fig. 5, and points up something else even more strongly, the possibilities of tremendous overshoot inherent in the logarithmic smoothing circuit. For $K=0.01$ a logarithmic slope is obtained that is four times the true growth factor, and this false reading is maintained for almost two periods. Still smaller K's will produce still larger overshoots, but a properly designed reactor control system will avoid this situation.

A similar set of calculations has been carried out for transients with negative periods. The results show that the logarithmic smoothing circuit responds well to negative periods at high flux levels and that it responds very inadequately to decreasing transients at low flux levels. So far as the design of reactor control systems is concerned, however, this seems to be an unimportant weakness.

RESPONSE TO RANDOM PULSES

Random Noise in the Linear Equivalent Circuit.

The simplest way to consider the random effects of neutron detections upon the logarithmic-diode smoothing circuit is to study the linear equivalent of that circuit shown in Fig. 2(C). In this way the variance of the output voltage can be estimated but not, of course, its mean, since Fig. 2(C) does not give a correct representation for the direct currents that flow.

The pulses supplied by the pulse-forming network to the logarithmic smoothing circuit must be very short, of the order of a microsecond in length, if excessively many counts are not to be lost at high counting rates. Since the time constant of the smoothing circuit must be much longer than this, the pulses may be regarded as mathematical impulses. The response of the linear approximate circuit to an input impulse carrying charge Q is, for $t\geq0$,

$$h_1(t)=(Q/C)\exp(-t/R_1C)=(Q/C)\exp(-nQt/AC) \qquad (31)$$

To the extent that the linear-network model applies, e_1 may be regarded as a shot noise composed of pulses having the shape given by equation 31, and the statistical methods of references 6, 7, and 9 may be applied. For counting rate n the autocorrelation function of the fluctuations in e_1 is found to be

$$\phi_{ee}(\tau)=(AQ/2C)\exp(-nQ|\tau|/AC) \qquad (32)$$

The magnitude of the fluctuations, as given by the variance, may be obtained from equation 32 by setting τ equal to zero. This gives

$$\mathrm{var}\{e_1\}=\phi_{ee}(0)=AQ/2C \qquad (33)$$

It is interesting to note that the magnitude of the fluctuations is independent of the counting rate. Because of this fact the logarithmic-diode counting-rate meter estimates all counting rates with the same percentage accuracy.

The power density spectrum of the fluctuations in e_1 is given by the Fourier transform of the autocorrelation function. The frequency bandwidth occupied by e_1 is seen to be proportional to the counting rate. Because the fluctuations become increasingly rapid at higher counting rates, without increasing in amplitude (by equation 33), the low-pass filter in the differentiating network becomes increasingly effective with larger n (see Fig. 3).

From shot-noise theory it may also be predicted that the random variable e_1, as well as e_2 and γ, is nearly normally distributed in amplitude.[7]

General Nonlinear Solution for Input Impulses

The input current i in the general solution of equation 21 need not be continuous; all that is required is that $i(t)$ be integrable. If $q(t)$ is the integral of the input current

$$q(t)=\int_0^t i(w)dw \tag{34}$$

equation 21 may be rewritten

$$z(t)=\{z(0)+(1/AC)\int_0^t \exp[q(w)/AC]dw\}\times \exp[-q(t)/AC] \tag{35}$$

If the input current contains impulses, this simply implies step discontinuities in $q(t)$ and $z(t)$.

Suppose that during the time interval $(0,t)$ the input consisted of M impulses, each of which carried a charge Q. For this type of input the integrand of equation 35 remains constant except for step discontinuities at the series of M epochs marked by the detection of neutrons. The integral of equation 35 thus becomes a finite sum, and the equation may be rewritten

$$z(t)=z(0)\exp(-MQ/AC)+ (1/AC)\sum_{m=0}^{M} t_m \exp(-mQ/AC) \tag{36}$$

where, in changed notation, t_0 denotes the length of time between time t and the most recent impulse, t_1 denotes the length of time between the next-to-last and last impulses, etc. Equation 36 gives $z(t)$ for any series of input impulses each of charge Q.

Nonlinear Solution for Constant Counting Rate

If the input impulses come from neutron detections of constant counting rate, they comprise a stationary shot noise. To say that a random process is stationary implies that the time origin may be chosen anywhere without affecting the statistical properties of the process. If the time origin of equation 36 is chosen in the very remote past, M approaches infinity, and in the limit

$$z=(1/AC)\sum_{m=0}^{\infty} t_m(\exp(-mQ/AC) \tag{37}$$

Since the input pulses occur at random, the t_m are random variables distributed in some fashion in amplitude.

Equation 37 gives a good deal of insight into the working of the logarithmic-diode counting-rate meter. The variable z, which is the antilogarithm of the output voltage, is made up of the weighted sum of the time intervals between pulses. The conventional counting-rate meter, which in basic form consists of a linear resistance and capacitance in parallel, estimates counting rate in quite a different manner. It assigns a weight proportional to $\exp(-t/RC)$ to each input impulse, where t is the elapsed time since the impulse occurred. If past time is divided into very small intervals of length Δt, if the number of counts that occurred in the last interval is denoted by n_0, and the number in the next most recent interval by n_1, and so on, the output voltage E is given by

$$E(t)=(Q/C)\sum_{m=0}^{\infty} n_m \exp(-m\Delta t/RC) \tag{38}$$

This method of estimating counting rate produces a weighted average of the number of input impulses that have occurred during recent time.

The contrast between these two methods represents a fundamental difference between two philosophies of counting-rate measurement. A counter can be used to count the number of pulses that occur in a predetermined interval of time, or it can measure the length of time required for a predetermined number of pulses to occur. The conventional and logarithmic-diode counting-rate meters conform roughly to the "fixed-time" and "fixed-count" philosophies respectively, but differ from counters in that they assign some weight to all past data. This inability to "forget" completely the remote past makes for sluggish response to transients in counting rate. Of the two types of counting-rate meters, however, the logarithmic-diode meter is the better where increasing counting rates are expected, for it makes possible more rapid response at the higher counting rates, where more information is available.

The author has been able to derive a number of the properties of the random variable z in equation 37, and have calculated, for special cases, the probability distribution of the logarithm of z, a variable that is linearly related to the output voltage of the logarithmic-diode counting-rate meter. The calculations are too lengthy to present here, but the results may be summarized as follows:

1. Mean diode voltage associated with input impulses of counting rate n and charge Q is slightly less than the voltage produced by a direct current of magnitude nQ. For a smoothing circuit with $AC/Q=1.44$ it is about 15% less, and for one with $AC/Q=2.89$ it is about 8% less. Since AC/Q will be 10 or more in a meter with adequate smoothing, it does not seem necessary to calibrate for this discrepancy.
2. The calculations for variance based on the linear approximate circuit (equation 33) produce results that are 6% and 3% too great respectively, for the two smoothing circuits defined in the foregoing.
3. The variance of e_1 is independent of the counting rate. A similar result was shown for the linear approximate circuit (equation 33).
4. The amplitude distribution of e_1, even for the small values of AC/Q just listed, is very nearly normal.

In short, for engineering purposes the linear approximate to the smoothing circuit is entirely adequate to calculate noise fluctuations.

Response to Increasing Counting Rate

If the logarithmic-diode smoothing circuit has been operating long enough for many counts to occur, equation 37 expresses its output, even if the input impulses have been generated during a counting-rate transient. If z is finite, it follows that only a finite number of t_m have contributed appreciably to the sum. If, during the total interval of time spanned by those t_m, the counting rate was nearly constant, the results of the foregoing sections apply directly. If, on the other hand, the counting rate increased appreciably during that interval of time, it follows that the earlier t_m (those with higher index m, therefore farther in the past) are on the average longer than the most recent ones. Hence z is larger than it would have been if the most recent counting rate had existed for a long time, and the indicated counting rate is low.

Consider how this qualitative reasoning applies in a particular situation. Imagine a counting-rate meter that observes about 40 pulses to estimate a constant counting rate, and assume that a long-established counting rate of 1/sec begins to increase with a period of 10 sec. After one period has passed, the counting rate will be less than 3/sec, the t_m that contribute appreciably will certainly include time before the transient began, and z will indicate much less than the true instantaneous counting rate. After seven periods have passed, however, the counting rate will exceed 1,000/sec, and 40 pulses will occupy only about 40 milliseconds on the average, which is only 0.004 period. For this higher counting rate the smoothing circuit responds as to a stationary input simply because its shorter memory time (here only about 40 milliseconds) does not span an appreciable change in counting rate.

Only when the input counting rate is apparently constant over the smoothing time of the logarithmic circuit does z approach a correct counting-rate indication. When a positive period is applied, the output of the meter at first lags the input; but eventually it catches up, and equation 33 is applicable for noise calculations. The duration of the smoothing-circuit transient may be estimated from Fig. 5, and random fluctuations during this interval are of no great interest.

Effect of Counting-Rate Transients on Period Meter

When an input with a finite period is applied to the whole period-measuring system, and after the smoothing circuit has begun to follow accurately the increasing counting rate, the differentiating network receives a ramp input plus noise of magnitude given by equation 33 and of frequency bandwidth that increases with n. If the reponse time of the differentiating network is much shorter than the period, the steady-state analysis given in an earlier section applies. For shorter periods, if the range of counting rates traversed during the response time of the differentiating networks is considered, the steady-state analysis gives upper and lower bounds for output noise to be expected when measuring nonzero growth factors.

Design of Low-Level Period Meter

EXTERNAL CONSTRAINTS

By far the most important constraint placed upon this type of period meter is the upper limit on counting rate achievable with a fission counter. The best counters have a resolving time of a few microseconds, and the resulting loss of counts must be considered at counting rates greater than about 20,000/sec. While some method of compensating for the lost counts may make it possible to operate at rates up to 100,000/sec, there is still a fundamental limit to the information input to the period meter. This limit sets a lower bound on the periods that can be measured, since shorter periods allow fewer total counts before the maximum counting rate is reached.

The designer is restricted, too, by the limited number of available logarithmic circuit elements, even though the analysis presented here applies to any 2-terminal logarithmic device. To the extent to which a logarithmic triode behaves like a logarithmic diode followed by a voltage amplifier, it too may be analyzed by the methods presented here.

When the logarithmic element is chosen, both A and Q are fixed. The former is a parameter of the logarithmic device; the latter is fixed because at maximum counting rate the logarithmic element should just saturate.

DESIGN PARAMETERS

After A and Q are settled, there remain only two fundamental parameters for the designer to choose. These are C, the smoothing capacitance, and T_2, the time constant of the differentiating circuit (the type of differentiating circuit introduced earlier is assumed for this discussion).

Choice of C and T_2 can best be made after consideration of two other system parameters. The first of these is AC/Q, a dimensionless parameter that expresses the average number of counts that occur during one "time constant" of the logarithmic circuit. This parameter indicates the amount of smoothing that takes place in that circuit; it occurs in many of the equations in the foregoing analysis.

The second parameter may be called the "breakpoint counting rate." It is that counting rate at which the time constant of the logarithmic smoothing circuit, given by AC/nQ, is equal to T_2. The breakpoint counting rate produces the noise maximum in the solid-line curves of Fig. 3. Its importance as a system parameter, however, stems from an observation generalized from the study of a number of alternate period meters: the transient response of the logarithmic smoothing circuit is so sluggish at low counting rates that no more than about one decade of useful range exists below the breakpoint counting rate.

If residual flux in the shutdown reactor is not sufficient to provide a counting-rate of one tenth the breakpoint rate, it may be advisable to provide a small bias current to the smoothing circuit to keep the voltage on C from going too far negative, and to prevent an input transient with an extremely small value of K; see Fig. 5.

If short periods must be reliably detected, then both C and T_2 must be kept small; if accuracy (low noise) is demanded, then response time must be sacrificed. Increasing C raises the breakpoint counting rate and lowers noise at the low counting rates; increasing T_2 decreases noise at all counting rates, especially at the high rates, and increases the minimum period that can be detected.

ACHIEVABLE PERFORMANCE

Values of AC/Q between about 10 and 30, and values of T_2 ranging between 2 and 5 sec produce period meters adequate over a useful range. Such meters can measure periods as short as 3 or 4 sec if those periods are initiated at the lower counting rates, and periods as long as several hundred seconds will be detected within the range of the meter, for noise becomes quite low at high counting rates. Period-measuring instruments of this type could profitably be used in a feedback loop to control start-up periods between, say, 10 and 20 sec.

The foregoing paragraphs give only an indication of a solution to a specific design problem. It is hoped that this paper will stimulate further use of logarithmic-diode counting-rate and period meters by making effective some improvement in their design.

References

1. CONTROL AND OPERATION OF A PILE, **W. J. Ozeroff.** "The Science and Engineering of Nuclear Power, Vol. 1," **Clark Goodman,** editor. Addison-Wesley Press, Cambridge, Mass., 1947, pp. 251–72.

2. AN INSTRUMENT FOR MEASURING THE LOGARITHM OF NEUTRON LEVEL AND THE PERIOD OF A PILE, **W. H. Jordan, H. B. Frey, G. Kelley.** *ORNL-110,* Oak Ridge National Laboratory, Oak Ridge, Tenn., 1948.

3. **R. E. Bell, R. L. Graham.** *Review of Scientific Instruments,* New York, N. Y., vol. 23, 1952, p. 301

4. INSTRUMENTATION AND CONTROL OF REACTORS, **J. D. Trimmer, W. H. Jordan.** *Nucleonics,* New York, N. Y., vol. 9, Oct. 1951, p. 60.

5. THE DECILOG: A UNIT FOR LOGARITHMIC MEASUREMENT, **E. I. Green.** *Electrical Engineering,* vol. 73, July 1954, pp. 597–99.

6. **S. O. Rice.** *Bell System Technical Journal,* New York, N. Y., vol. 23, 1944, pp. 282–332; vol. 24, 1945, pp. 46–156.

7. **David Middleton.** *Journal of Applied Physics,* New York, N. Y., vol. 22, 1951, pp. 1143–63.

8. **W. M. Grim, Jr., B. B. Barrow, J. C. Simons, Jr.** *Nuclear Science and Engineering,* Academic Press, New York, N. Y., vol. 1, 1956, pp. 80–91.

9. **W. R. Bennett.** *Proceedings,* Institute of Radio Engineers, New York, N. Y., vol. 44, 1956, pp. 609–38.

10. HIGHER TRANSCENDENTAL FUNCTIONS, VOL. II (book), **Arthur Erdélyi, et al.** McGraw-Hill Book Company, Inc., New York, N. Y., 1953, pp. 143–45.

11. TABLES OF SINE, COSINE, AND EXPONENTIAL INTEGRALS, VOLS. I, II. National Bureau of Standards, Washington, D. C., 1940

A Wire-Activation Technique for Reactor-Flux-Profile Measurements

ALTON E. KLICKMAN
NONMEMBER AIEE

FRANCIS R. DEFALCO
NONMEMBER AIEE

Synopsis: The application of the wire-activation technique was extended so that flux profiles could be obtained in the cores of pressurized water-type reactors. Tests with magnesium with 0.01 a/o (% by atoms) indium, 0.010 inch in diameter, showed that the method is feasible and the counting rates are high enough to give reliable results.

THE method of determining a neutron flux from measurements made on materials which have been activated by the flux has been used successfully by many people. The materials most commonly used are gold and indium foils. The activation technique has been used to measure the flux profiles in several new reactor cores.

In the reactor core under consideration, a measurement of the flux profile along a line parallel to the vertical axis of the core was desired. To measure this profile, a vertical well of a long tube of Zircaloy 2 extending from the top of the reactor vessel to the bottom of the core was placed in the reactor. The inside diameter of this well is 0.210 inch and the total length is 20 feet. The reactor was of the pressurized water type, so that the measurements had to be made at high temperatures and in high fluxes. It would be difficult to position foils in the well accurately and since it would take a long time to make enough foil measurements to obtain a complete flux profile, a technique using the wire form of the material to be activated appeared to be a solution. A length of wire can be positioned in the core accurately and enough data for a complete flux profile can be obtained quickly. In the wire-activation technique as used, a cadmium-alloy tubing surrounding the wire is used to simulate the cadmium cover used in the foil-activation technique.

Theory

Briefly, the procedure used in the activation technique consists of exposing a foil of the material being used as a detector to the neutron flux for a known period of time. The sample is removed and its activity measured. If the procedure is repeated with a foil wrapped with cadmium, the foil will be activated only by neutrons with energies above the cadmium cutoff, about 0.4 ev (electron volt). Hence, the resulting activity is less and is related to the neutron flux above 0.4 ev.

The difference between the number of counts obtained from the two foils can be related to the thermal flux, or to the thermal-neutron density if the detector cross section is inversely proportional to neutron velocity. The ratio of the activities is called the cadmium ratio and includes the ratio of thermal to resonance flux.

Specifically, the detector with a thermal-activation cross section σ_{th} has a response to a thermal-neutron flux $(nv)_{th}$ given by equation 1.

$$\text{thermal response } \alpha (nv)_{th}\sigma_{th} \tag{1}$$

and a response to the resonance flux given by equation 2

$$\text{resonance response } \alpha \int_0^{u_{co}} \sigma_a(u)\phi(u)du \tag{2}$$

where

u = lethargy corresponding to neutron energy $E = \ln 10^7 \text{ ev}/E(\text{ev})$
u_{co} = lethargy corresponding to cadmium cut-off energy
$\phi(u)$ = flux per unit lethargy at u

Some materials have one sharp resonance peak in the cross section, and effectively all the epithermal neutrons absorbed by the material are neutrons which have energies near the resonance value. In this case, one may write the resonance response as

$$\text{resonance response } \alpha \phi(u_r)\int_0^{u_{co}} \sigma_a(u)du \tag{3}$$

where u_r is the lethargy corresponding to the energy of the resonance peak, since the integral is primarily determined by the energy interval in the neighborhood of the resonance peak.

From the definition of the cadmium ratio the following result is obtained

$$R_{cd}-1=\frac{\text{thermal response}}{\text{resonance response}} = \frac{(nv)_{th}\sigma_{th}}{\phi(u_r)\int_0^{u_{co}}\sigma_a(u)du} \tag{4}$$

R_{cd} = cadmium ratio

Thus the actual cadmium ratio observed is a function of the sensitivity of the detector to resonance and thermal neutrons and does not lead to the true flux ratio unless this sensitivity is known. However, if the values for the thermal cross section and the resonance-activation integral are known, the ratio of the fluxes can be obtained.

The thermal response is obtained in practice as the difference between the uncovered exposure (measuring total response) and the cadmium-covered exposure (measuring resonance response). If some standard is available, this thermal response can be converted to absolute thermal flux. Then equation 4 can be used to obtain the resonance flux per unit lethargy if the value of the integral is known.

By activating materials which have prominent resonance absorption at selected energies, it is possible to measure the flux spectrum over a considerable range of energies. Several detector materials, e.g., indium, cobalt, and manganese, were investigated in this program, with the main effort being devoted to indium. The treatment discussed in the foregoing is valid for indium because it has one sharp resonance peak at 1.44 ev.

Apparatus and Procedures

The objectives of this program were twofold. First, a wire material had to be developed for use as the measuring device. To make this a useful scheme, a system for irradiating the wire and removing it after irradiation was needed. Second, a system had to be developed for scanning, counting, and recording the counts.

Description of Equipment

The device developed in this program for measuring spatial distribution of neutron flux consisted of a magnesium-0.01 a/o indium wire strung inside a tube of either magnesium or magnesium with 33 a/o cadmium. The wire and tube were attached to a long aluminum handling rod. The length of the wire and tube section were chosen to exceed the length of the reactor core by several feet. The aluminum handling rod was long enough to extend through the shield, a

---

Paper 57-405, recommended by the AIEE Nucleonic and Radiation Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE 2nd Nuclear Engineering and Science Conference, Philadelphia, Pa., March 11–14, 1957. Manuscript submitted April 16, 1957; made available for printing June 19, 1957.

ALTON E. KLICKMAN is with Battelle Memorial Institute, Columbus, Ohio, and FRANCIS R. DEFALCO is with Westinghouse Electric Corporation, Pittsburgh, Pa.

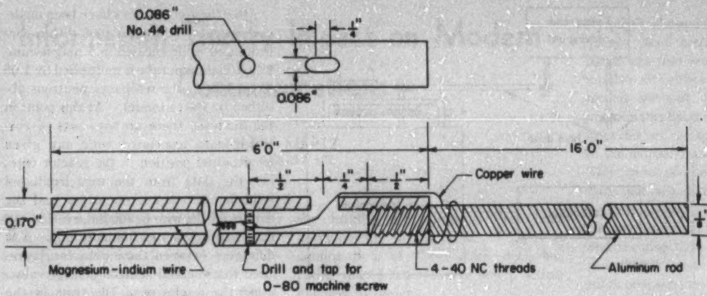

0.086"
No. 44 drill
$\frac{1}{4}$"
0.086"
6'0"
16'0"
$\frac{1}{2}$"
$\frac{1}{4}$"
$\frac{1}{2}$"
Copper wire
0.170"
$\frac{1}{8}$"
Magnesium-indium wire
Drill and tap for
0-80 machine screw
4-40 NC threads
Aluminum rod

Fig. 1. Flux-measuring device

distance of some 20 feet. A disk was attached to the handling rod to locate the wire-tube assembly axially in the reactor core well and to actuate an electric timer. The arrangement of the assembled device in the core well of the device is shown schematically in Fig. 1.

To make the wire usable in high-temperature surroundings, the indium required alloying. The amount of indium was chosen to give a high enough counting rate, a low gamma heating, and a low activity for easier handling. Magnesium, chosen as the base, is a light material and a good diluent from the standpoint of gamma heating. In addition, it has only a short-lived activity. The half life of indium 116 is 54 minutes and the half life of magnesium 27 is only 9.6 minutes.

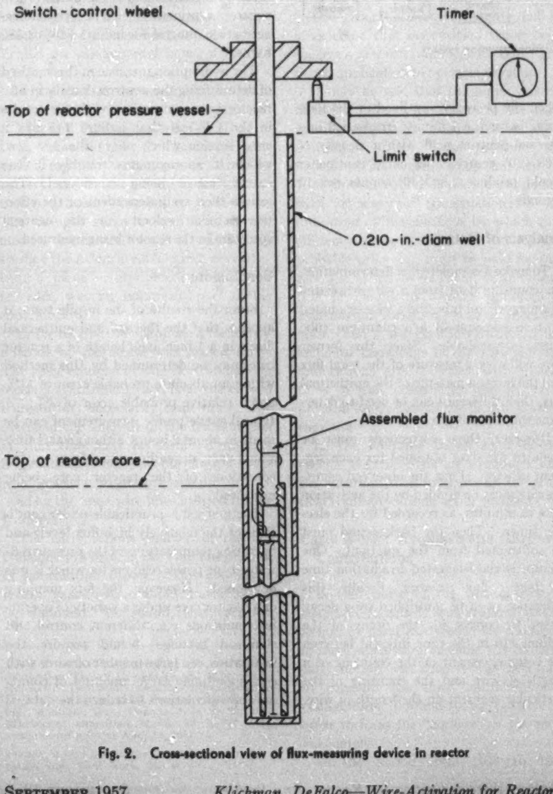

Switch-control wheel
Timer
Top of reactor pressure vessel
Limit switch
0.210-in.-diam well
Assembled flux monitor
Top of reactor core

Fig. 2. Cross-sectional view of flux-measuring device in reactor

The second major component developed in this program, the scanning bench for measuring the activation of the wire after irradiation, consists of two lead-shielded reel mechanisms, a lead collimator, a scintillation counter, and a timing-and-driving unit. Block diagrams of these units are shown in Figs. 2 and 3. These were designed so that the activated wire can be wound onto one reel and then counted automatically as it passes from one reel to the second reel by the collimator-counter assembly. The timing-and-driving mechanism moves the wire intermittently at the rate of 1 inch per 15 seconds. At the end of each 15-second period the number of counts recorded during the period is printed out automatically the scaler resets, and the wire advances 1 inch.

The scintillation counter was a shielded sodium iodide (thallium-activated) crystal with a Dumont *6292* photomultiplier operated at 1,250 volts: see Fig. 4. The power supply was an Atomic superstable high-voltage power supply and the preamplifier, amplifier, and single-channel analyzer were supplied by Detectolab. The digital scanner and recorder were Berkeley models *1556-1* and *1450*, respectively. These were modified so that one signal from the timing-and-driving unit would control both the printing operation of the recorder and the resetting operation of the scanner. Because a shielded sodium iodide crystal was used, only gamma rays were counted. It was found experimentally that satisfactory counting rates could be obtained if all gamma rays above a very low energy were counted, discriminating only against noise pulses.

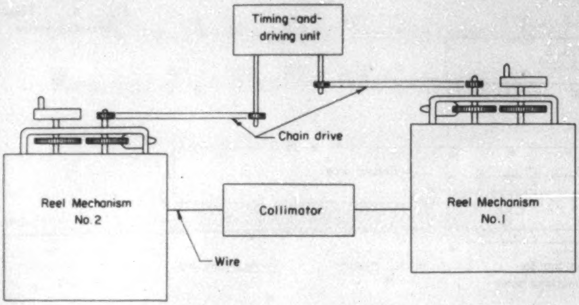


Fig. 3. Plan view of scanning bench

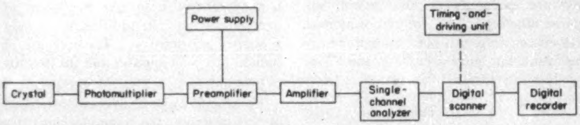


Fig. 4. Block diagram of scintillation counting system

A single-channel analyzer is used to permit the spectrum analysis necessary for the periodic recalibration of the scanning bench with standard sources.

Description of Operating Procedure

The operating procedure developed for this equipment is as follows:

1. The wire-tube assembly is placed in the reactor core well and irradiated for a period of 1 minute. The actual time is a function of the neutron density and varies with the reactor.
2. The wire-tube assembly is removed from the core, the wire is removed from the tube assembly, and the wire is wound into one of the reel mechanisms.
3. One hour after removal from the reactor, the wire is processed through the scanning bench and the activity is measured. The 1-hour period is observed to allow the decay of extraneous activities.

Results of In-Pile Tests

To calibrate the wire materials that were considered for use as the flux-measuring device, in-pile tests were run at the CP-5 at Argonne. In addition, these tests were used to develop techniques for activating and counting the wire.

In these in-pile tests it was found that when the procedure outlined in the foregoing is used, a 1-minute irradiation in a thermal-neutron field with a density of 4.38×10^8 neutrons per cubic centimeter would produce 2.30×10^5 counts per 15 seconds.

Analysis of Data

To make a complete flux determination, the counting data from a wire irradiated in a magnesium tube and a wire irradiated in a magnesium-33 a/o cadmium tube must be available. Since the former wire will give a measure of the total flux and the latter a measure of the epithermal flux, their difference can be used to determine the thermal flux.

However, three corrections must be made to the data obtained for each segment of wire. First the observed counting rate must be divided by the activation time in minutes, as recorded by the electric timer. Then the background must be subtracted from the quotient. One minute is the suggested irradiation time for higher flux reactors. Finally, this difference must be multiplied by a decay factor to correct for the decay of the indium 116 in the time interval between the commencement of the counting of a length of wire and the counting of the particular segment on the length of wire.

After these corrections have been made, each piece of data associated with a wire irradiated while encased in a magnesium-33 a/o cadmium tube is multiplied by 1.05 to correct for the resonance neutrons absorbed by the cadmium. At this point in the analysis, there are two sets of corrected data associated with any given longitudinal position in the reactor core; i.e., the data from the wire irradiated while in the magnesium tubing and the data from the wire irradiated while in the magnesium-33 a/o cadmium tubing. The difference between these data can be related to a thermal-neutron density value, using the results of in-pile tests in the CP-5. The ratio of these data, the cadmium ratio, is related to the thermal flux and the resonance flux per unit lethargy by equation 4.

If the thermal flux is defined as the product of the thermal-neutron density and a neutron velocity, i.e., 2,200 meters per second, the proper activation cross section for indium 115 is 145 barns (10^{-24} square centimeters). The resonance-activation integral for indium 115 is 2,580 barns.

The assumption intrinsic in this method of determining the neutron density in any reactor by a comparison with the results in the CP-5 is that indium 115 has a cross section which obeys the 1/v law, which it approximates roughly in the energy range being considered. The results then are independent of the effective neutron velocity or the neutron spectrum in the reactor being measured.

Conclusions

From the results of the in-pile tests, it appears that the thermal and epithermal fluxes in a 1-inch axial length of a reactor core may be determined by this method within an absolute probable error of 11% and a relative probable error of 3%. A typical single-profile measurement can be made in about 2 hours, although this time would vary, depending upon the operating conditions of the reactor core being measured.

The method is practicable under conditions of the relatively high-flux levels and operating temperatures of the pressurized-water-type power reactors for which it was developed. However, the flux mapping of a reactor core under a variety of operating conditions, e.g., different control and shim-rod settings, would require the irradiation of a large number of wires with correspondingly large amounts of counting and calculations to reduce the data.

Information Theory Impact on Modern Communications

PIERRE MERTZ
NONMEMBER AIEE

THERE ARE TWO ASPECTS to the development of communications systems (as there are to the development of many other systems). The first is the invention of the fundamental equipment. The second aspect is the analytical study of the use to be made of this equipment, namely, of the transmission medium in which the equipments are to work, and the engineering of the component parts to give an optimum result.

Much of the design of communications transmission systems has been based on intuitional ideas. This resulted in part from the fact that these systems were mostly based on terminal devices which formed the fundamental bases of the inventions.

In the case of the telegraph the basic equipment was of course the transmitter or key, and the tape recorder or sounder. Later came the relay. These were connected together with a power source or battery, and at first merely a wire with return to ground between the two stations. The thought given to transmission was mostly intuitional until Lord Kelvin studied the Atlantic cable[1] (and, nevertheless, a piece of equipment, the siphon recorder, was an important part of his contribution).

In the case of the telephone a frequency analysis of the signal was sensed early. The basic equipment, the electromagnetic receiver-transmitter (and later the carbon transmitter) necessarily came first. The status of the analysis of line transmission at that time may be surmised when it is recalled that one of the burning questions was whether insulators should not be used on the open-wire lines utilized for the transmission![2]

The early serious analysis of the line transmission became the foundation stone of information theory. This art has in recent years had a rapid mathematical development. So sound, however, was the intuitional optimizing that later developments are to a large degree a confirmation of the earlier decisions, rather than a generation of new practical engineering designs.

One notices that in the literature there is a large gap between mathematical papers that treat of elaborate methods to refine the selection of a signal received under unfavorable transmission conditions, and descriptive papers that treat of actual communication systems designed to operate with maximum effectiveness by using methods of this kind.

This general status is changing, and it is expected that information theory will suggest a new art, and methods that have actual application to operating systems.

It may be said that the major contribution of information theory to the communications art is a mathematical systemization of the broad processes involved. This systemization has led to 1. an intensification of the preoccupation with efficiency of transmission; 2. an indication of the methods by which this efficiency may be increased and of the probable price that must be paid for this increase, as well as its value; 3. an indication of means to combat unfavorable transmission conditions such as excessive noise; and 4. a logical integration of the technical knowledge in the various communications arts, together with indications of the possibilities for a better physical integration of the facilities which they use.

In the following pages some of the specific contributions are examined, and the nature of their influence on the practical communications art is outlined.

Early Work: Nyquist's and Hartley's Contributions

In the course of the transmission of telegraph signals it was realized that a frequency band was needed for the signal in the transmission medium. A frequent early notion was that it was necessary to transmit the third harmonic of a combination dot-and-space pattern in order to form the "corners" of the wave adequately.

In the period from 1924 to 1928 Nyquist[3,4] showed that the minimum frequency band needed for the telegraph signal was that required to send a given number of independent quantities per second without ambiguity, and that mathematically the frequency bandwidth in cycles per second was half the number of the independent quantities per second. The figure of one-half comes from the fact that the Fourier component at a given frequency has two parameters, say, the amplitude and the phase (or the cosine and the sine terms). The "independent quantities" referred to are, in telegraph, mark or space indications. The choice of only two possible values for each quantity is, however, very restricted. The theory is general enough to treat of the distinction of one of a range of values for each independent quantity.

Nyquist showed that the characteristic distortion or intersymbol interference or, again, interference between the signal for one independent quantity and that for another, could be reduced to zero by an exact shaping of the electrical transfer characteristic, specified as a function of frequency. He showed that the independent quantities could be measured in several ways on the received signal, and that the required shape of transfer characteristic depended upon which way was chosen. Among the several ways or criteria of measuring, one was the amplitude of the signal wave at the exact middle of the signal element duration. Another criterion was the exact spacing between times of transition between various critical amplitudes. Still another criterion was the integrated impulse of the wave amplitude with respect to time during the signal element.

Nyquist recognized that there is difficulty in cutting off a frequency band sharply to the minimum width needed. He indicated various ways of shaping a "roll-off" about the cutoff frequency without reintroducing the interference between the independent quantities. These roll-offs have again to be shaped differently according to the specific criterion chosen for measuring the independent quantities.

Nyquist also indicated that a practical beneficial effect of adding this roll-off was to render more lenient the exactness of the line-up of the system while still keeping a low value of intersymbol interference. The roll-off used a somewhat wider band in the frequency spectrum than would a sharp cutoff, and this served as payment in return for the beneficial effect.

Nyquist's disclosures made it possible to design rationally the frequency band-

---

Paper 57-635, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 20, 1957; made available for printing April 16, 1957.

PIERRE MERTZ is with the Bell Telephone Laboratories, Inc., New York, N. Y.

width which was needed to transmit a telegraph signal, or perhaps, by its more modern name, a data transmission signal. It also gave good grounds for estimating the band that would be required for picture transmission and television signals, although in these cases the ultimate receiver is the eye, and it may not follow the rules assumed for the decoding relays or the equivalent of the telegraph or data system. Some discussion of these is given later.

Paralleling the work of Nyquist were the studies of Hartley[5] and of Küpfmüller.[6] Hartley's contribution may probably best be expressed in his own statement of what has come to be known as Hartley's law:

> The total amount of information which may be transmitted over a system (whose transmission is limited to frequencies lying in a restricted range) is proportional to the product of the frequency-range which it transmits by the time during which it is available for the transmission.

Noise Filtering: Wiener's Contributions

In the measurement of a physical quantity, small errors are likely to be made. As the quantity varies and the measurements continue, these erratic errors superpose, in the measurements, on the actual changes in the physical quantity. A complete record of the measurements therefore shows a more erratic, or less "smooth," behavior than the actual quantity itself.

This is quite analogous to what is called the "noise" which is superposed on a transmitted signal; i.e., the variations of the noisy signal include both the variations of the original signal and the random variations of the noise. This is often "smoothed," in the case of the data, by using running averages; in the case of electrical signals by "filtering" them through a low-pass or band-pass filter. There is a danger in such processes in overdoing the smoothing, to the point of suppressing, at least in part, the actual variations of the physical or original quantity. Or if the smoothing is insufficient, not as much of the noise is suppressed as could be under the circumstances.

Norbert Wiener,[7] in the period 1942–1949, indicated a set of criteria by which an optimum filter could be designed. Essentially these criteria depend upon an advance knowledge of the correlation among signal amplitudes in successive instants of time and, similarly, also a correlation among noise amplitudes. The most effective filter can obviously be obtained where the signal correlation is high (or extends over appreciable time intervals) and the noise correlation low (or extends over only short time intervals). Essentially this of course means that the signal frequency band is narrow and the noise band wide. The qualitative conclusion here is effectively that which would be obtained from earlier practice. The Wiener contribution however gives exact methods for designing an optimum filter, which methods also hold under less clear-cut qualitative circumstances.

This suggestion of Wiener for the design of an optimum filter has initiated a great wealth of additional suggestions[8–14] and refinements. The objective of the Wiener criterion is to reduce, as far as is possible, the deviation of the actual filter output from what would have been the signal output without noise. This deviation comprises both the change in the signal alone caused by the filter, and the residual noise itself, after reduction by the filter. The additional suggestions are often in the direction of changing this criterion of optimization. Also a considerable fraction of them are aimed at the problem of recovering returned radar signals, either from a great range so that they are immersed in strong noise or under otherwise extremely unfavorable conditions. In this work one finds the names of Zadeh, Ragazzini, Lee, Chang, North, Woodward, Davies, Middleton, Fano, Booton, Slattery, Bose, di Toro, and Bendat, to mention only a very few.

While the contributions of these authors are great, and not to be minimized, their primary advances have been of a conceptual nature and few of them have been applied to actual communication links. Three conditions have generally conspired to lead to this result. These are the following.

First, the mathematics for optimizing a filter are complicated, and except for very simple conditions, become prohibitive. The mathematics of Wiener are formidable enough, but in the case of most of the other authors they are frequently much worse.

Second, the equipment which is required to implement most of these ideas tends easily to become prohibitive in bulk and cost.

Third, the advantages gained by applying successive refinements to this problem become progressively smaller. A comparatively simple filter, say, roughly approximating the Wiener design, gives in practice a large proportion of the advantage to be expected. Further elaborations are hardly justified in general practical circumstances.

The Measurement of Information: Gabor's Terms

The first step negotiated in information theory is really to develop a quantitative notion of what constitutes information. There are a number of aspects to this concept and several of them are examined here.

According to a first concept (paralleling the ideas of Nyquist and Hartley, but elaborated by Gabor,[15,16]) information is that which imparts a representation to a receptor (eventually a person) and that the more complex the pattern of the representation, the more information is required.

In this concept the representation is structurally separable into distinguishable groups or categories, i.e., the representation can be described by a catalog list of items. As an illustration, the representation may be that of a solid figure, say the form of a washer. Here the catalog list is very short, and comprises outside diameter, inside diameter, and thickness. The representation may be more generally that of a polynomial, in which each item represents the coefficient of a given term. Or the representation may be in other terms, but in each case it appears as a catalogued list of items.

The term "logon" has been given by Gabor to indicate this structural complexity: one logon is that which enables one new distinguishable group or category to be added to a representation.

Once the categories of a representation are listed, it is necessary to complete the representation by specifying a quantitative factor for each. This requires a scale. Unit increments along this scale measure the precision of the measurement, on a scale of probable errors (50% of readings for a given quantity within the increment, 50% outside). Such unit increments have been termed "metrons," and the scale a "proper scale."

In this concept the logons are obviously the "independent quantities," of Nyquist. The metrons are simply the quantitative values attached to the independent quantities.

The quantitative notion of information in later years has been generalized in terms of order and disorder concepts. According to this idea, lack of information on an array presumes the possibility of a large number of such arrays from which to choose. This is conceived as a situation of disorder. As information is obtained on the array the choice from the large number narrows down to a choice from a smaller number, and more order appears. This view becomes analogous to thermodynamic concepts, say, of the states of a gas. The quantitative measure of order as against disorder in thermodynamics is the entropy. This (with a negative sign attached, which will not be discussed here) has therefore been proposed as a quantitative measure of information by Shannon[17] and Szilard.[18] The concept will be considered in somewhat more detail in the next section.

More sophisticated measures of the amount of information have been recently proposed in which the criterion is based upon the unexpectedness and importance of the information, or upon concepts obtained from game theory.[19,20] In one of the latter the measure of the information is based on possible profits of a gambler betting on the subject matter of the information.

These can be of conceptual interest in developing broad philosophy, and can have later impact on communication systems. They would not have a very immediate effect on the design of systems intended to take any information impartially and send it accurately and impassively to the receiver. They would however, give a possible clue regarding tolerances on errors in the transmission.

Measurement of Information in Terms of Entropy: Shannon's Contributions

In 1948 Shannon[17] generalized Nyquist's results into a comprehensive theory. It is of course impossible to present here anything but a sketchy notion of the drift of these contributions, and the original must be consulted for a more comprehensive view. As has been noted, Shannon defined quantity of information in a fashion analogous to entropy in thermodynamics. Specifically, with this approach the average information H per symbol, or entropy in a coded sequence is

$$H=-\Sigma_j P_j \log_2 P_j \qquad (1)$$

where P_j = probability of jth configuration or symbol and where the summation is taken over all possible configurations, and where the probabilities of each of these are independent of the others.

A simple example of such a sequence might be the number 37621. This consists of five successive symbols. Each of these represents some one of the ten configurations of the ten Arabic numerals. In their general use in numbers like this the probabilities are independent, equal, and each equal to 1/10. In such a case

$$H=-10\times(1/10)\log_2(1/10)$$
$$=3.32$$

When information of the kind described by equation 1 is transmitted over a communication channel, some errors will in general be experienced. The entropy of the delivered signals is then less than that of the input signals as given by equation 1.

In particular the signals may be translated into variations of voltage which are impressed upon a circuit. The frequency band of this circuit may be limited to a specific width. The circuit may have superposed noise, which is responsible for the errors which have been assumed. The exact steps of the reasoning cannot be covered here, but it develops that where the noise is suitably random, the entropy of the received impaired signals is equal to the entropy of the unimpaired signals less the entropy of the noise.

$$R=H(s)-H(n) \qquad (2)$$

When this is further developed, the maximum transmission capacity (or information per second) of the band-limited noisy channel is

$$R_{\max}=W\log_2\left(\frac{S}{N}+1\right) \qquad (3)$$

where

W = frequency bandwidth of channel in cycles per second
S = average power limitation of signal
N = average power of noise

This quantity $R_{\max}$ represents the maximum number of bits per second which can be sent through the channel under the power and frequency bandwidth conditions considered, with a possibility of a vanishingly small percentage of error. This vanishingly small percentage of error assumes an optimum coding, which in general requires merging large groups of symbols of a nonoptimum code, and representing these by large but different groups of the optimum code.

A simple illustration may bring out the significance of this law. Consider a system transmitting two binary digits per second per cycle of frequency band W. Such would be a very common system. By Shannon's equation (equation 3)

$$\log_2\left(\frac{S}{N}+1\right)=2 \qquad (4)$$

$$\frac{S}{N}+1=4 \qquad (5)$$

$$\frac{S}{N}=3=5\text{ db (decibels)} \qquad (6)$$

Actually in a simple conventional practical system using signal ON for marking and OFF for spacing, noise peaks 6 db down from marking level (or therefore 3 db down from average signal power level) cause errors when of unfavorable polarity. With Gaussian noise, the probability of this is one in 10^5 when the peaks are 13 db above rms. Consequently a signal-to-noise ratio of $13+3=16$ db, average signal power to average noise power, is needed to maintain an error rate not greater than one in 10^6 bits.

Comparison with Shannon's equation shows a penalty of 11 db for not using the optimum code. The first obvious change would, of course, be to a polar rather than a neutral code, i.e., one would use + amplitude for a mark, and −amplitude for a space. This saves 3 db and still leaves an 8-db penalty.

From this point on the nature of the preferred codes requires a longer explanation. It may be assumed that the message appears in the conventional binary form, and that it is desired to use it also in this form at the distant end to which the information is to be transmitted. Essentially, then, a translation to the preferred code is necessary at the transmitting end of the communication facility, and an inverse translation is necessary at the receiving end. The translator at each end includes such storage as may be necessary to give the required freedom for the construction of the preferred code to be used over the channel path.

It is noted by Shannon that the greatest entropy for the signal is obtained when it has a Gaussian amplitude distribution. He therefore suggests for an optimum code that the signal information be grouped into messages containing specific numbers of bits. For each different possible message a signal is constructed, which has a Gaussian amplitude distribution, and therefore looks like a sample of random noise. The detection is carried out by comparing the received wave with the constructed signals for all possible messages. That message is selected whose signal shows the least rms discrepancy from the received wave. As the message lengths are increased, this type of coding and reception can approach a recovery of the 8-db residual disadvantage just noted.

Rice[21] has explored two possible mechanisms for the approximation of such signals with Gaussian distributions. In the one mechanism the signals are constructed of Fourier polynomials, i.e., finite sums of harmonically related sine and cosine components which are therefore uniformly spaced in the allotted frequency interval. The coefficients of the terms are chosen at random from a universe of coefficients that give a fixed average signal power for each constructed signal. In the second mechanism the signals are constructed of $(\sin x)/x$ components. The arguments x are functions of time and the peaks (for $x=0$) are evenly spaced at successive instants of time through the message duration. The components have amplitude coefficients chosen at random from a universe of coefficients that gives a fixed average signal power for each constructed signal.

With each of these constructed signals the ideal rate of signaling (as given by equation 3) is approached, for any arbitrarily small assigned percentage of error, as the message length chosen in forming the code is increased. This means effectively that the 8-db penalty which has been mentioned can be recovered, at the expense of storing and translating sufficiently long message groups to bring the error frequency to not more than the one in 10^6 bits assumed immediately following equation 6.

The advantage obtained by such codes comes in two steps. First, there is the advantage obtained from the use of a Gaussian distribution of amplitudes, to give the signal a maximum entropy. Second, there is the advantage obtained from a long coding sequence. The exact nature of the latter advantage is somewhat more elusive. Essentially, however, the idea is to take advantage of the exponential growth in number of choices selected by the signal with the number of bits of quantity of information represented in the signal. This exponential is the inverse of the logarithm in equation 1. The number of active choices is the element that remains constant in the transition from one side to the other of the translators. However, by providing a large number of additional and unused choices in the channel side of the translator, redundance can be built into the signal in this location. It is in this location that the major sources of error are found. The ratio of redundant to utilized choices can be made very large for a modest increase in number of bits represented if the coding sequence is long enough, because of the exponential relation between them.

A simple illustration of this that does not, however, involve the signal with Gaussian distribution, can start with a group of two binary digits that represents four choices. At a 50% increase in speed three digits in the same time will represent eight choices. Thus each of the four choices in the original code can be represented by two choices in the speeded-up code.

However, if the original code group is taken as six (instead of two) digits, it represents 64 choices. With a $33^1/_3$% increase in speed (i.e., less than the previously considered 50%) eight digits can be sent in the same time. These represent 256 choices, or a possibility of 4 choices for each of the original code. The freedom to do this increases rapidly as the code group is lengthened.

It has been shown by Shannon,[22] Elias,[23] and others that, if the utilized choices are located judiciously among the much larger number of redundant choices, errors in the utilized choices will lead most often to surrounding redundant choices and only much more rarely to other utilized choices.

The term "surrounding" refers to the possibility, which will not be further detailed here, of plotting all the choices in a multidimensional hyperspace. In this hyperspace plot, choices represented by very nearly the same code designations cluster into groups.

A suitable receiver can translate the surrounding choices that have been referred to into the "surrounded" utilized choice, which erases the error.

This arrangement can greatly reduce the frequency of occurrence of errors on the system. It requires, of course, some measure of slowing down of the actual rate of useful transmission as evaluated by equation 3. In the limit, as the code groups are taken infinitely long, the amount of slowing down from the ideal rate becomes infinitesimal.

Another type of signal from those of Rice has been suggested by Golay.[24] Here the information is carried by the timing of a short duration pulse of larger instantaneous amplitude (but low average power over a given signaling time). The exact instant at which the pulse occurs, over this given time, indicates which one of a finite group of messages is that transmitted. It appears that the use of this code permits approaching the efficiency indicated by equation 3.

Slepian[25] has studied concrete codes in binary form to obtain the advantages of large groupings. These, of course, do not refer to the band-limited portion of the channel.

While the elaboration of codes in this manner has proved of great conceptual value, it has not, so far, generally found favor in practical communications. Among the reasons is that basically, as Shannon points out, it introduces delay for encoding and decoding. Further, it involves translating, and possibly the acquisition of specialized receiving equipment which is bulky and costly and which may not be in economic equlibrium with the transmission facility, and the possibility of additional error hazards is introduced. Needless to say, these points are all under careful scrutiny at present, and it is not unlikely that in time some derivation of the ideas will find a place in practical utilization.

Perhaps it should be observed that, even more fundamentally, the Shannon equation 3 really represents an upper limit of performance for idealized limitations only of frequency band and noise. The simple equation is not to be used where it does not apply, and other derivations are applicable where other limitations are significantly present.

The imperfect equalization of the line, and variations with time of its net loss, are additional limitations which encroach on the upper limit of performance, on a basis of economic equilibrium. Also the noise is not usually of the idealized type, the frequency band does not cut off sharply (as noted earlier by Nyquist), and the timing of sampling instants at the receiver differs from optimum.

This all means that a performance well under that indicated by the idealized Shannon equation is likely to be in economic equilibrium, and that the value of comparatively elusive highest efficiency codes to actual communications systems is restricted, at least in the present state of the art.

Some work has been done evaluating the influence of imperfect equalization and regulation.[26–29] These both increase the signal-to-noise requirement for a given error performance. A conventional figure for engineering the equalization is to permit the residual error to degrade the noise performance by 3 db. Tolerances on the additional error to be contributed by imperfect regulation vary, of course, but these also run in that general order of magnitude.

Carrier Transmission, Envelope Rectifier, and Homodyne Detector

In the discussions so far base-band transmission of the signals is assumed, i.e., it is assumed that the amplitude of a direct current or voltage has been modulated or adjusted according to a given or input wave shape to denote an instantaneous proportional value at each moment.

In fact, carrier transmission is more likely to be used in practice. It was noted by Carson[30–32] that amplitude modulation of the carrier leads to symmetrical side bands about the carrier, each of which is a replica of the base-band spectrum. The signal information is duplicated in the side bands so that a loss in efficiency results unless other steps are taken.

Such a step can consist of filtering out one side band, leading to a single sideband transmission. Or, alternatively, it can take the form of a separate and independent modulation of the sine and the cosine components of the carrier wave respectively. This gives two individual channels in the double side band, which recovers the information efficiency.

These steps are incidentally not available (at least in this clean-cut form) to frequency modulation. Thus the various advantages of such modulation are purchased at the price of a reduced information-carrying efficiency.

Transmission by amplitude-modulated carrier leads to an examination of detectors used to recover the signal. The simplest of such detectors is the envelope rectifier. This cannot recover negative excursions of the envelope as distinguished from positive excursions. When a carrier signal contains such negative envelope excursions it is said to be "overmodulated."

The "homodyne" detector[31] is a product demodulator which multiplies the incoming signal with the carefully phased output of a local oscillator of the exact carrier frequency, and integrates the result in a low-pass filter. It can distinguish between aiding and opposing phase in the incoming carrier as compared with the local carrier, and therefore can recover distinguishable positive and negative envelope excursions. The homodyne detector has the possibilities ascribed to an "ideal" detector as developed in some viewpoints of information theory,[33] when used with an optimum noise filter. It is the detector which has been assumed earlier as available to permit the use of a polar carrier signal as distinguished from a mere on-off signal.

The price of the homodyne detector operation is, of course, the necessity for the local carrier supply of exact frequency and phase. This can be supplied by some extra channel in the transmission, or by a periodic reference condition in the signal, during which the phase of a local oscillator can be adjusted. Because of this generally high price, some modifications have occasionally been proposed to the conventional homodyne detector for special signals.[34]

Also because of this high price, it is frequently found that economic equilibrium in the design of practical communications systems leads to the use of a simple envelope rectifier, in spite of its poorer signal-to-noise performance.

Timing Indications as Information

A point not always noted in the study of communications systems by information theory techniques is that timing indications are a form of information. This has been revealed by Golay,[24] and also in the information theory study of radar return signals.[35] In these, of course, the timing indication is vital; it measures the range of the object revealed, and the ability to do this accurately is one of the major characteristics of radar.

In actual point-to-point and network communications systems the timing may also be of importance. Some facilities are used for timing indications alone as, for example, in the radio transmission of time signals, or in the synchronization of astronomical or similar observations.

In telegraph or data transmission, however, timing is also of importance.[29] It is of sufficient importance in many cases for a special portion of the information capacity to be assigned to it. The timing is of importance because, as indicated by Nyquist, there is an optimum time for sampling a received signal impulse to determine its amplitude (or when in a binary code, whether it is marking or spacing).

This timing information can on a hypothetical basis be set during a lining-up period and need not be repeated. Or it may be derived from signal transition times, as in the Baudot quadruplex.[36] But, practically, it is found desirable to repeat this fairly frequently, as there may be short occasional circuit interruptions or bursts of noise that lead the receiver to lose count. This repetition permits the receiver to start afresh with a minimum of lost time.

The timing information is assuming particular importance in pulse code modulation (PCM) systems, to be described later, where regenerative repeating is used.[37,38] In such cases it is essential that the regeneration cover exactness in the timing of the regenerated pulses, as well as in their wave shape. It is also essential that whatever is used as timing channel also be regenerated in such a way that the timing accuracy is not too much impaired.

Applications of Information Theory to Pictorial Transmission, Facsimile, and Television

The objective of a fairly large group of communications systems is the transmission of 2-dimensional pictures. These include facsimile (both photographic and the cruder direct recording) and television.

In these systems the end result is a picture seen by the ultimate viewer. The principal requirement is really that his subjective needs be satisfied. These needs cover such things as perceptions of sharpness, tone rendering, contrast, freedom from extraneous patterns, and fringes in space or in time (the latter including, for example, flicker and stroboscopic and after-image effects), and (in color systems) color rendering.

Such subjective perceptions would not immediately be expected to be measurable in terms of information content. This is, of course, the actual situation, and in some instances the fact has been corroborated. Nevertheless it has been found that the techniques of information theory have been quite powerful in the engineering of transmission systems for pictorial communications.

In particular it has been found that the logon-carrying capacity of the system (i.e., the number of independent picture luminance values per second that can be transmitted) as measured by the frequency bandwidth available, correlates reasonably well with the resolving power and sharpness of the received image. It is clear that since the optical profession has been searching for many years for an objective measure of sharpness, and has succeeded in getting only a rather loose correlation,[39–41] that the correlation for facsimile and television cannot be expected to be any more precise. However, at least a reasonably workable measure of this aspect of picture quality has been obtainable from the information theory approach.[42–46]

This has been so successful that the optical profession has itself begun to try the approach in some of its problems of optical degradation.[47–49] In this technique the subjective degradation is identified with the drop in reproduced information caused by the impairment.

The National Television Systems Committee standards on a color television broadcast signal, adopted by the Federal Communications Commission, has also made use of conclusions from information theory.[50] It was long known that the composition of a conventional television signal is such as to make rather inefficient use of the frequency space used for the information transmitted.[51] Even at an earlier date, Gray suggested the interleaving of three channels for color transmission.[52]

Later it was further discovered that the eye, in the resolving of image detail, appreciates chromatic information less than it does luminance information. It even needs less chromatic information for certain color groupings than for others. Thus it needs less logon content in the signal, or a lower frequency band, for a purely chromatic than for a luminance signal, and less for one of the chromatic signals than for the other (if the two are suitably chosen.)

In the National Television Systems' Committee signal the two chromatic channels are much narrower than the luminance channel, and one chromatic channel is narrower than the other. The two chromatic channels are then multiplexed on the luminance channel by interleaving in the same frequency space.

In this manner the same frequency space as used for a monochrome television signal is made to carry more information for the color television signal. The extra information has been tailored as far as possible to the subjective needs of the ultimate viewer, and inserted in inefficiently used portions of the original signal.

In recent times, serious proposals have been made for reducing the inefficiency of transmission of monochrome television.[53] This of course has been an objective for many years, but the results, until very lately, were disappointing. Even the system here proposed is usable only where the signal has been translated from its normal or analog form into a digital form. The proposal consists in separating the high-frequency from the low-frequency components of the signal and translating them into digits separately. The low-frequency digits, which occur relatively seldom, are translated with customary accuracy. The high-frequency digits, which occur at a high rate, are translated on a very coarse scale of gradations. In an experimental system this reduced the frequency band required for the digital transmission to about half.

Applications of Information Theory to Telephony

The application of information theory to telephony again brings in the subjective factors that have been noted with regard to pictorial transmission.

This matter involves the problems of the intelligibility and entropy in linguistic processes. Studies have been made of the point both for written and auditory languages. There are, of course, two basic steps in coding an idea into a telephone line signal. The first is to express the idea in terms of a language. The second is to translate the language into line signals. In the speech form, the language consists of a succession of elementary sounds which are called "phonemes."

It is not expected that in telephony the speaker will modify his speech to increase his efficiency or reduce the redundance in the first step. It is, however, of interest to know the amount of the redundance because it gives a notion as to how much the signal may be distorted in transmission without destroying the sense conveyed by the speech.

Shannon[54] has made some estimates of this for printed English text of 100 characters, on the basis of asking a subject to guess the successive letters, and noting the statistics of correct answers. This redundancy emerges, in his terms, in the order of 75%. Fry[55] has carried out this test with spoken English, counting phonemes instead of letters, resulting in a redundancy of the order of 55%.

With regard to the second step, one may note that according to the speed of the speech, one may utter from 2 to 20 phonemes per second, with a possible "typical" value of 8. The number of phonemes, say, in English, varies according to point of view. If one aims primarily at intelligibility and considers closely similar sounds as one phoneme, the number perhaps is in the fifties and each can be identified by a 6-bit signal. If one aims at fine distinctions, the number exceeds 100, and, say, seven bits are necessary. This means that merely intelligible speech requires some 12 to 120 bits per second of basic information. A typical value might be 50. If finer distinctions of speech are considered, the numbers still run only from 14 to 140. A little more should also be added for average pitch indication, and possibly for other indications characterizing expression.

Experiments have been carried out to determine the various fundamental transmission requirements for a telephone circuit to transmit speech in the conventional manner, and with a degree of ease deemed suitable for message communication.[56,57] These lead specifically to frequency band needs of the order of 3,000 cycles.

A 3,000-cycle band is very inefficiently used by a process which conveys only some 15 to 150 bits per second of information, especially when more than half of that is redundant to the ideas expressed. The fact illustrates, however, why it is possible to carry out the speech transmission under very unfavorable conditions and still retain a useful degree of intelligibility. It also illustrates why it is more difficult to use a telephone facility for data transmission than for speech transmission (even aside, of course, from the fact that it was designed for speech in the first place).

Because of these considerations there has been an effort in a number of directions to transmit speech in a narrower frequency band than used by a conventional telephone circuit. Essentially the problem results from the fact that the code which is used to represent the voice, namely the analog connection between the signal voltage and the instantaneous air pressure (or velocity) in the region between the speaker's mouth and the microphone, is not an efficient code. It is necessary here to beware of any efficient code development that depends upon a long code group for its efficacy. For if too much encoding and decoding time are to be required, this can quickly impede the to-and-fro flow of conversation. Such a flow is one of the outstanding advantages of telephony over other media of long-distance communication, and its possibility is of great importance to the art.

Of the various efforts to improve the coding efficiency of speech, the one that has probably been pushed the furthest toward a practical realization is the "vocoder."[58]

The objective of the vocoder is to correlate the speech to some degree with the mechanism of the voice. In one form of the scheme the 3,000-cycle speech spectrum is divided into 10 bands, and the energy in each noted. This is communicated to the far end in channels that have a 25-cycle base band (or 50-cycle double side-band width.) An 11th channel, of the same width, indicates the fundamental pitch. At the receiver these signals control the intensities (in the original bands) of a sound stream consisting of a hiss or a buzz (the choice of these two being under control of the pitch channel). The entire speech channel is therefore reduced to some $11 \times 50 = 550$ cycles, and could be reduced further by vestigial side-band transmission through the individual channels if necessary.

The net contributions which information theory has made to conventional telephony are thus an awareness of the price paid in information efficiency for the simplicity that characterizes the ordinary message telephone system, and the means which can be used to increase this efficiency in cases where the economic equilibrium warrants it.

Telephony in addition to this also comprises switching and signaling circuits that have much in common with the binary code pulses of telegraphy and data transmission. In these circuits the contributions of information theory run along much the same paths as for these two other arts.

In recent times there has been much consideration of the possibilities in a conversion of the speech signal itself into a binary digital code. (This is called "pulse code modulation," PCM.[59]) When the signal is in such a code form it much resembles a digital data signal. The principal differences are, first, that interruptions for synchronizing or other auxiliary signals, where necessary, must be kept down to very small intervals of time so as not to interrupt the free flow of speech and, second, that the criteria for permissible error rates are different. Aside from these differences, all of the objectives hold for the improvement of data transmission by the application of information theory. The further advantage over data transmission is that the facility obtained is not just a by-product of a much more widely used facility, and the new plant therefore can be specifically tailored to the PCM transmission.

References

1. On the Theory of the Electric Telegraph, Mathematical and Physical Papers, Vol. 2 (book), **Lord Kelvin.** Cambridge University Press, Cambridge, England, 1884.

2. Conferences Are an Old Bell System Custom, **R. E. Mooney.** *Bell Telephone Magazine*, New York, N. Y., vol. 34, Winter 1955–56, pp. 259–68.

3. Certain Factors Affecting Telegraph Speed, **H. Nyquist.** *Bell System Technical Journal*, New York, N. Y., vol. 3, Apr. 1924, pp. 324–46.

4. Certain Topics in Telegraph Transmission Theory, **H. Nyquist.** *AIEE Transactions*, vol. 47, Apr. 1928, pp. 617–44.

5. Transmission of Information, **R. V. L. Hartley.** *Bell System Technical Journal*, vol. 7, July 1928, pp. 535–63.

6. Transient Phenomena in Wave Filters, **K. Küpfmüller.** *Elektrische Nachrichten-Technik*, Berlin, Germany, vol. 1, 1924, pp. 141–52.

7. Extrapolation, Interpolation, and Smoothing of Stationary Time Series (book), **N. Wiener.** John Wiley & Sons, Inc., New York, N. Y., 1949.

8. Optimum Filters for the Detection of Signals in Noise, **L. A. Zadeh, J. R. Ragazzini.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 40, Oct. 1952, pp. 1223–31.

9. An Extension of Wiener's Theory of Prediction, **L. A. Zadeh, J. R. Ragazzini.** *Journal of Applied Physics*, New York, N. Y., vol. 21, July 1950, pp. 645–55.

10. Information Theory and Inverse Probability in Telecommunication, **P. M. Woodward, I. L. Davies.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. III, Mar. 1952, pp. 37–44.

11. On Determining the Presence of Signals in Noise, **I. L. Davies.** *Ibid.*, vol. 99, pt. III, Mar. 1952, pp. 45–51.

12. Analysis of the Factors Which Determine Signal-to-Noise Discrimination in Radar, **D. O. North.** *Report PTR-6C*, RCA Laboratories, Princeton, N. J., June 1943.

13. An Optimized Theory for Time-Varying Linear Systems with Nonstationary Statistical Inputs, **R. C. Booton.** *Proceedings*, Institute of Radio Engineers, vol. 40, Aug. 1952, pp. 977–81.

14. Statistical Theory of Signal Detection, **D. Middleton.** *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, vol. 3, Mar. 1954, pp. 26–51.

15. Theory of Communication, **D. Gabor.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 93, pt. III, Nov. 1946, pp. 429–57.

16. The Nomenclature of Information Theory, **D. M. MacKay.** *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, vol. 1, Feb. 1953, pp. 9–21.

17. A Mathematical Theory of Communication, **C. E. Shannon.** *Bell System Technical Journal*, vol. 27, July 1948, pp. 379–423; Oct. 1948, pp. 623–56.

18. On the Entropy Reduction in a Thermodynamic System by Interventions of Intelligence, **L. Szilard.** *Zeitschrift für Physik*, Berlin-Charlottenburg, Germany, vol. 35, 1929, pp. 840–56.

19. Quantal Aspects of Scientific Information, **D. M. MacKay.** *Transactions*, Professional Group on Information Theory-1, Institute of Radio Engineers, Feb. 1953, pp. 60–80.

20. A New Interpretation of Information Rate, **J. L. Kelly.** *Bell System Technical Journal*, vol. 35, July 1956, pp. 917–26.

21. Communication in the Presence of Noise—Probability of Error for Two Encoding Schemes, **S. O. Rice.** *Bell System Technical Journal*, vol. 29, Jan. 1950, pp. 60–93.

22. Communication in the Presence of Noise, **C. E. Shannon.** *Proceedings*, Institute of Radio Engineers, vol. 37, Jan. 1949, pp. 10–21.

23. Coding for Noisy Channels, **P. Elias.** *Convention Record, Ibid.*, pt. 4, Mar. 1955, pp. 37–46.

24. Note on Theoretical Efficiency of Information Reception with PPM, **M. J. E. Golay.** *Proceedings*, Institute of Radio Engineers, vol. 37, Sept. 1949, p. 1031.

25. A Class of Binary Signaling Alphabets, **D. Slepian.** *Bell System Technical Journal*, vol. 35, Jan. 1956, pp. 203–34.

26. Theoretical Fundamentals of Pulse Transmission, **E. D. Sunde.** *Ibid.*, vol. 33, May 1954, pp. 721–88; July 1954, pp. 987–1010.

27. Effects of Signal Fluctuation on the Detection of Pulse Signals in Noise, **M. Schwartz.** *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, vol. 2, June 1956, pp. 66–71.

28. Communication Through Noisy, Random-Multipath Channels, **G. L. Turin.** *Convention Record*, Institute of Radio Engineers, pt. 5, Mar. 1956, pp. 154–66.

29. Transmission Line Characteristics and Effects on Pulse Transmission, **P. Mertz.** *Symposium on Information Networks*, Brooklyn Polytechnic Institute, Brooklyn, N. Y., vol. 3, Apr. 1954, pp. 85–114.

30. Signaling with High Frequency Waves. *U. S. Patent no. 1,449,382*, **J. R. Carson.** Mar. 27, 1923 (filed Dec. 1, 1915).

31. Relations of Carrier and Sidebands in Radio Transmission, **R. V. L. Hartley.** *Proceedings*, Institute of Radio Engineers, vol. 11, Feb. 1923, pp. 34–56.

32. Early History of Single Sideband Transmission, **A. A. Oswald.** *Ibid.*, vol. 44, Dec. 1956, pp. 1676–79. (The remainder of this issue is also devoted to single sideband transmission.)

33. The Theory of Signal Detectability, **W. W. Peterson, T. G. Birdsall, W. C. Fox.** *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, vol. 4, pp. 171–212, Sept. 1954.

34. Predicted Wave Radio Teleprinter, **M. L. Doelz.** *Electronics*, New York, N. Y., vol. 27, Dec. 1954, pp. 166–69.

35. Theory of Radar Information, **P. M. Woodward.** *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, vol. 1, Feb. 1953, pp. 108–13.

36. Coding and Information, la Cybernétique, **E. Picault.** "Editions de la Revue d'Optique Théorique et Instrumentale," **L. de Broglie,** editor. Paris, France, 1951, pp. 297–314.

37. Transistorized Binary Pulse Regenerator, **L. R. Wrathall.** *Bell System Technical Journal*, vol. 35, Sept. 1956, pp. 1059–84.

38. Transistor Pulse Regenerative Amplifiers, **F. S. Tendick.** *Ibid.*, pp. 1085–1114.

39. Resolving Power Criteria and the Intensity Epoch Slope, **B. P. Ramsay, E. L. Cleveland, O. T. Koppins.** *Journal*, Optical Society of America, New York, N. Y., vol. 31, pp. 26–33, Jan. 1941.

40. The Nature and Evaluation of the Sharpness of Photographic Images, **G. C. Higgins, L. A. Jones.** *Journal*, Society of Motion Picture and Television Engineers, New York, N. Y., vol. 58, Apr. 1952, pp. 277–90.

41. Relation of Definition to Sharpness and Resolving Power in a Photographic System, **G. C. Higgins, R. N. Wolfe.** *Journal*, Optical Society of America, vol. 45, Feb. 1955, pp. 121–29.

42. The Production and Utilization of Television Signals, **F. Gray, J. W. Horton, R. C. Mathes.** *Bell System Technical Journal*, vol. 6, Oct. 1927, pp. 560–603.

43. Wire Transmission System for Television, **D. K. Gannett, E. I. Green.** *Ibid.*, pp. 616–32.

44. Quality Rating of Television Images, **P. Mertz, A. D. Fowler, H. N. Christopher.** *Proceedings*, Institute of Radio Engineers, vol. 38, Nov. 1950, pp. 1269–83.

45. The Evaluation of Picture Quality with Special Reference to Television Systems, **L. C. Jesty, N. R. Phelp.** *Marconi Review*, London, England, vol. 14, July 1951, pp. 113–36; Oct. 1951, 156–86.

46. Image Gradation, Graininess, and Sharpness in Television and Motion Picture Systems, **O. H. Schade.** *Journal*, Society of Motion Picture and Television Engineers, vol. 56, Feb. 1951; pp. 137–77; vol. 58, Mar. 1952, pp. 181–222; vol. 61, Aug. 1953, pp. 97–164; vol. 64, Nov. 1955, pp. 593–617.

47. Resolving Power and Information, **G. Toraldo di Francia.** *Journal*, Optical Society of America, vol. 45, July 1955, pp. 497–501.

48. Diffraction and Quantity of Information, **A. Blanc-La Pierre, M. Perrot.** *Comptes Rendus*, Institut de France, Academie des Sciences, Paris, France, vol. 231, 1950, pp. 539–41.

49. On the Assessment of Optical Images, **P. B. Fellgett, E. H. Linfoot.** *Philosophical Transactions*, Royal Society of London, London, England, vol. 247, Feb. 1955, pp. 369–407.

50. Principles of Color Television (book), **K. McIlwain, C. E. Dean.** John Wiley & Sons, Inc., New York, N. Y., 1956.

51. A Theory of Scanning, **P. Mertz, F. Gray.** *Bell System Technical Journal*, vol. 13, July 1934, pp. 464–515.

52. Electro-Optical Transmission System. *U. S. Patent no. 1,769,920*, **F. Gray.** July 8, 1930 (filed Apr. 30, 1929).

53. Reduced-Alphabet Representation of Television Signals, **E. R. Kretzmer.** *Convention Record*, Institute of Radio Engineers, pt. 4, Mar. 1956, pp. 140–47.

54. The Prediction and Entropy of Printed English, **C. E. Shannon.** *Bell System Technical Journal*, vol. 30, Jan. 1951, pp. 50–64.

55. Communication Theory and Linguistic Theory, **D. B. Fry.** *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, vol. 1, Feb. 1953, pp. 120–24. (Presented at the London Symposium on Information Theory, Sept. 1950.)

56. Speech and Hearing (book), **H. Fletcher.** D. Van Nostrand Company, Inc., Princeton, N. J., 1929.

57. The Words and Sounds of Telephone Conversations, **N. R. French, C. W. Carter, W. Koenig.** *Bell System Technical Journal*, vol. 9, Apr. 1930, pp. 290–324.

58. Remaking Speech, **H. W. Dudley.** *Journal*, Acoustical Society of America, New York, N. Y., vol. 11, Oct. 1939, pp. 169–77.

59. The Philosophy of PCM, **B. M. Oliver, J. R. Pierce, C. E. Shannon.** *Proceedings*, Institute of Radio Engineers, vol. 36, Nov. 1948, pp. 1324–31.

An Improved Single-Phase Watt-Hour Meter

J. G. LANDRY
NONMEMBER AIEE

J. F. SCAMMAN
ASSOCIATE MEMBER AIEE

THE PURPOSE of this paper is to present the design considerations of a new single-phase or single-stator watt-hour meter with marked improvement in performance, and lower inherent manufacturing costs than a conventional meter of the same heavy load capacity. A limited discussion of design problems on present ratings is included to clarify the comparisons made.

Until fairly recently, the modern self-contained watt-hour meter has been manufactured in only two nominal current ratings. The 15-ampere (amp) meter, with a maximum capacity of 100 amp, is used where accuracy at relatively light loads is important and is becoming known as a class *100* meter. The 50-amp meter with a maximum capacity of 200 amp, is used where relatively heavy loads are encountered and it is becoming known as a class *200* meter.

Both of these conventional meters have essentially a one-piece combined voltage and current electromagnet. The voltage coils, consisting of many turns of fine wire, are essentially the same, but the current coils differ considerably. For performance and economic reasons, it is desirable to represent each of the two circuits of a 3-wire meter on each of the two current lamination poles and to limit the maximum ampere-turns to 400. This is accomplished on the class *100* meter by having each line make a turn first around one pole and then around the other. Thus, at the maximum load of 100 amp, the four turns of the two current circuits yield 400 ampere-turns. With the class *200* meter each of the two circuits is divided into two branches, with each branch carrying only half the maximum 200 ampere current and encircling only one of the two current poles. Thus again the maximum ampere-turns equal 400.

Obviously the divided coils of the class *200* meter are more complex and hence more costly than the class *100* construction, and in addition there are other disadvantages. For instance, if manufacturing control is such that equal division of the load current is not accomplished, error torques which affect metering accuracy are developed. Furthermore, the two branches of a line, being joined at their extremities, constitute a closed loop or short-circuited turn enclosing both current poles. This loop cannot help but be energized by the useful potential coil flux. This results in increased potential circuit or "standby" losses and reduced meter torque. The phase angle of the working potential flux and the inherent temperature errors are also affected.

Despite the performance drawbacks of the conventional class *200* watt-hour meter, particularly at light loads, the continuing steady increase in the use of electric energy is resulting in greater usage of this rating of meter. Upon review of the design of this meter it became apparent to the authors that really significant performance improvement could not be achieved because of the inherent characteristics of its stator.

A completely new approach was undertaken which resulted in, first of all, a simple single-turn current coil for each of the two current circuits in the meter. These new coils are shown at the left of Fig. 1, and one of the styles of the conventional coils is shown at the right. This new construction completely eliminates all of the undesirable side effects such as higher losses, torque depression, and other effects previously mentioned. This is because these new coils do not form a closed or short-circuited loop. Each coil consists of a single wrap or turn about the laminations and therefore cannot be a source of losses. Also the two single turns encircle the current laminations at a point midway between the two current poles and they are farther from and normal to the disc. Both of these factors greatly improve the balance between the individual coils and reduce the sensitivity to coil position. The conventional meter, having its coils parallel to and close to the disc may exhibit calibration differences between individual coil excitation of about 1%. The new coils, however, exhibit calibration differences of less than 0.2% when similarly tested. Because of its simplicity the inherent cost is reduced and the new coil can be threaded through instead of slipped over the current poles. This new assembly technique permits the spacing between the current pole tips to be determined on an optimum performance basis, rather than on a mechanical clearance basis.

A second important improvement has been obtained by magnetically separating the current and potential electromagnets. This greatly reduces the intermingling of current and potential fluxes and the two electromagnets become more responsive to changes in the values they are each required to measure. Fig. 2 shows the lamination structure of the conventional combined electromagnet stator. For simplicity the current flux and potential flux distributions are shown on separate outlines. Mentally superimposing the two outlines the areas of intermingling become obvious.

Generally speaking the intermingling in the potential lamination areas of this combined core affect accuracy at low power factors, while the intermingling in the current areas has as its major effect poor performance at light loads. Let us consider, for example, the current electromagnet, composed of the coils or windings and the lamination structure. In the conventional stator the current lamination structure is required to carry the flux resulting from current in the current windings. This flux provides the disc torque component which should be representative of the load current or amperes passing through the current coils of the

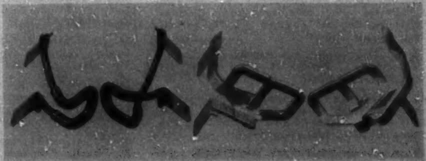


Fig. 1. Comparison of current coils

Paper 57-651, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 27, 1957; made available for printing April 7, 1957.

J. G. LANDRY and J. F. SCAMMAN are with the General Electric Company, Somersworth, N. H.

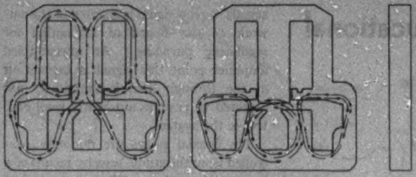


Fig. 2. Conventional stator laminations

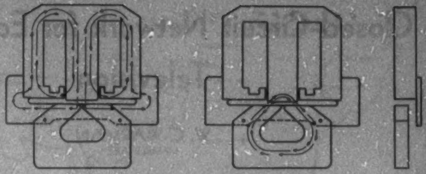


Fig. 3. New stator laminations

meter. In addition to this flux which may be considered a signal, the current lamination structure is required to carry, simultaneously, a much larger amount of potential flux through the greater portion of its effective path length. This potential flux, while useful in maintaining the permeability of the current lamination structure at a desirable level, causes the significance of the current flux, or signal, to be diminished. This is particularly true at low values of load current where the signal may be very small and represent only a few percent of the total flux present. The visible indications of this are inaccuracy at loads other than at which the meter is calibrated, and a variation in performance characteristics within a group of meters. A similar examination of the potential electromagnet of the conventional stator will show that a like situation exists.

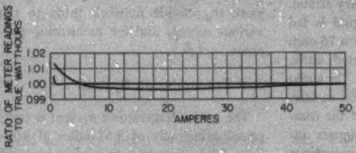


Fig. 4 (above left). Partial load curve

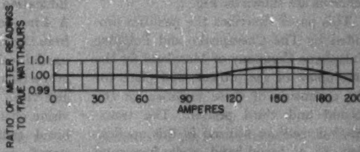


Fig. 5 (above right). Load curve

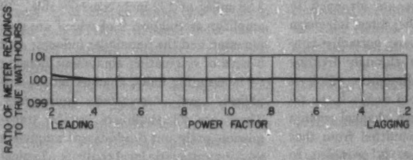


Fig. 6 (left). Power factor curve

It has been found that by magnetically separating the two electromagnets this intermingling of fluxes may be greatly reduced. Fig. 3 shows the lamination structure of the new stator. The side view indicates how the magnetic separation is accomplished. Again for simplicity the current and potential fluxes are shown on separate outlines. Mentally superimposing these two outlines, it becomes obvious that for all practical purposes, intermingling in the critical areas is eliminated. Thus the signal flux in each electromagnet becomes the primary flux present. This permits each of the electromagnets to more accurately indicate small changes in the values of the quantities they are measuring.

The single factor which has made possible these simplifications and improvements is the potential return lamination assembly. This is a lamination assembly which is fastened to the potential lamination assembly (or may be an integral part of several of the potential laminations) and has a portion adjacent to but spaced from the current lamination assembly. Thus there is provided a return path for useful potential flux. This novel lamination assembly permits the rearranging of the current coils into the simple form employed and permits the desired magnetic isolation of the stator electromagnets with no loss in torque-producing efficiency.

The effect of this new stator on the light load performance of the meter may be graphically shown (Fig. 4) by comparing the first 50 amp of its load curve with that of its class *200* predecessor. The elimination of detrimental flux intermingling is the primary reason that straight line accuracy is maintained even much below the 3-amp light load calibration point.

Examining the complete load curve of the new meter (Fig. 5), the excellent performance at heavy loads is made possible by the fact that the simplified current coil design permits the effective pole areas to be located for maximum efficiency, as previously mentioned. The accurate power factor curve (Fig. 6) for this new meter further shows the advantage of magnetically isolating the current and potential electromagnets to reduce flux intermingling.

A further major consideration is that of determining the rating of the meter. An evaluation of the inherent calibration linearity and low damping errors indicated the desirability of designing the meter for a full load test current rating of 30 amp. Specifically, the reduction of heavy load damping errors will now permit accurate metering at a higher rotor speed, and such a rating will provide maximum accuracy at light loads. This change in nominal rating from the historic 50 amp to 30 amp would normally be expected to pose new problems in rotor bearing design because of higher speeds and greater side thrust at maximum load. However, the magnetically suspended rotor bearing system of this meter is fully capable of operating under these conditions without experiencing any difficulties.

It is the opinion of the authors and others conversant with watt-hour meter design that this new meter incorporates the first significant changes in stator design in the past 30 years. Although this discussion has been limited to single stator watt-hour meters, the principles involved in the stator design are equally applicable to multistator watt-hour meters.

Closed-Circuit Network for Educational Television

W. C. WARMAN
MEMBER AIEE

THE Washington County, Maryland, Board of Education is conducting a 5 year experiment to evaluate the use of closed-circuit television as a teaching medium. The application of the medium is divided into three phases corresponding to the school years of 1956–57, 1957–58, and 1958–61 as shown in Fig. 1. The first phase provides for the interconnection of five program originating points with eight public schools in Hagerstown and vicinity with three television channels. The second phase provides for the expansion of the network to include the interconnection of 23 schools with 6 television channels. The third phase provides for the further expansion of the network to include all 48 schools in the County. The existing and planned extensions are shown on Fig. 1.

This paper describes the facilities provided by The Chesapeake and Potomac Telephone Company of Maryland for the first phase of the project, and plans for the expansion of these facilities for the second and third phases. The transmission medium utilized in this application is a broad-band radio-frequency transmission system.

The system is designed to accept video and audio signals at program originating points, convert these signals to standard television broadcast signals by means of miniature television transmitters, and transmit them over coaxial facilities to terminal locations. At the school terminals, connections are established to the master antenna distribution system, which in turn is connected to standard television receivers located in the classrooms. The television receivers may be tuned to the desired channel. Broadband amplifiers, capable of providing the ultimate requirement of six channels, are utilized in this system.

The facilities for the first phase are shown geographically on Fig. 1(A) and operate on a "round robin" basis, thus permitting programs to be inserted at various originating points and transmitted to the schools over the same line facilities. Arrangements provide for originating three simultaneous programs from the Board of Education studios and individual programs from studios located at North High School, the Library, the Museum, and South High School. Three programs are transmitted over the system as television channels 2, 4, and 6.

It may be of interest to note that the facilities for the first phase were planned, engineered, procured, installed, and aligned in approximately 2 months. The facilities are maintained by regular telephone maintenance men with electronic background and training secured during the installation and testing of facilities.

Network Description

The equipment component and line facilities for the Hagerstown system are shown schematically on Fig. 2. From an operational standpoint, the Board of Education location is the primary studio. A 4-megacycle (mc) video signal is fed from the studio equipment over a 75-ohm coaxial cable and the associated audio over a balanced 600-ohm pair. The audio and video signals are modulated to television channels 2, 4, and 6, and the combined radio-frequency (r-f) signals inserted on the network. Connections through coaxial relays are arranged so that the output of the three television channel transmitters are normally connected to the system through dividing networks. A band elimination filter is inserted in each channel to block the returning signal of a given channel when a program is being transmitted from that location. Amplification is provided to compensate for the cumulative line, pad, and filter losses.

In addition to the television channel signals, a pilot signal is inserted in the network at the Board of Education for regulating purposes. At every third amplifier point this signal is tapped off of the through line, is amplified, and detected to provide automatic level control in the system.

The coaxial relays are remotely controlled from the studio area by means of a key and lamp box. When it is desired to originate a program at any location other than the Board of Education, the television signals are introduced into the round robin by the operation of the appropriate coaxial relays at the remote location. Coincidentally, the channel-blocking filter is removed from the line at the Board of Education, thus permitting the transmission of the remotely originated program to pass through the Board of Education to the remaining schools on the network.

School locations which have no originating requirement and are not located near the round robin are served as branches off the network. Currently, there are two such schools but ultimately these branches will extend throughout Washington County. Directional coupling devices are used throughout the network to provide network feeds to the various schools and for monitoring purposes.

R-F Transmission System

The initial Hagerstown system is comprised principally of 8.57 miles of 0.92-inch solid dielectric triaxial cable and 3.88 miles of 1/2-inch coaxial cable, 19 amplifier assemblies, 9 of which are pole mounted and the remainder indoors; and 6 bays of originating equipment housing r-f modulating units, channel filters, amplifiers, and associated equipments. There are 23 amplifiers in the transmission path and 6 additional amplifiers used for automatic level control. Typical equipment installations are shown in Figs. 3 through 7. All line and equipment components are of a nominal 75 ohms impedance.

LINE FACILITIES

Two types of line facilities were utilized in this system, JEL 100 (Times Wire Manufacturing Company) and Spirafil coaxial cable (Phelps Dodge Copper Products Corporation). Both cable restrict r-f radiation to acceptable levels. The JEL 100 cable is a 0.92-inch solid polyethylene dielectric triaxial cable with a 9-gauge solid copper wire center conductor and two copper-braided outer shields insulated from each other by a layer of polyethylene insulation. The cable is covered

---

Paper 57-668, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 29, 1957; made available for printing May 7, 1957.

W. C. WARMAN is with The Chesapeake and Potomac Telephone Company, Baltimore, Md.

The author wishes to acknowledge the valuable assistance of the following people in the systems engineering of this project: W. Weller Holloway and J. M. Stack, The Chesapeake and Potomac Telephone Company of Maryland; Harry O. Saunders, of the American Telephone and Telegraph Company; J. M. Barstow and A. D. Fowler of the Bell Telephone Laboratories, Inc.; and Winfield W. Bemis and Lester C. Smith, Spencer-Kennedy Laboratories, Inc.

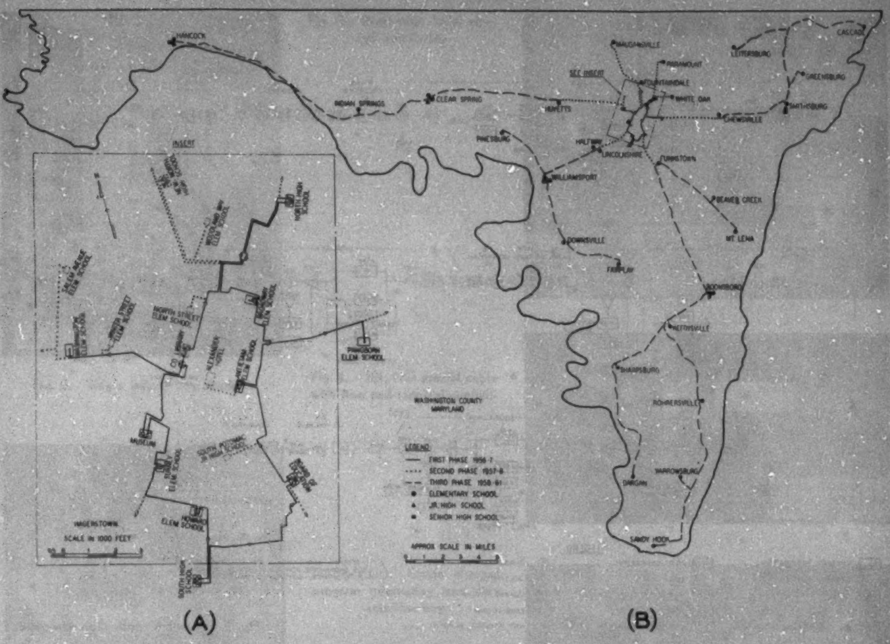

HANCOCK
INDIAN SPRINGS
CLEAR SPRING
HUYETTS
MAUGANSVILLE
PARAMOUNT
SEE INSERT
FOUNTAINDALE
WHITE OAK
LEITERSBURG
CASCADE
GREENSBURG
SMITHSBURG
CHEWSVILLE
PINESBURG
HALFWAY
LINCOLNSHIRE
FUNKSTOWN
WILLIAMSPORT
BEAVER CREEK
MT LENA
DOWNSVILLE
FAIRPLAY
BOONSBORO
KEEDYSVILLE
SHARPSBURG
ROHRERSVILLE
BROWNSVILLE
DARGAN
SANDY HOOK
INSERT
NORTH HIGH SCHOOL
NORTH STREET ELEM SCHOOL
PANGBORN ELEM SCHOOL
MUSEUM
SOUTH POTOMAC JR HIGH SCHOOL
SOUTH HIGH SCHOOL
HAGERSTOWN
SCALE IN 1000 FEET
WASHINGTON COUNTY MARYLAND
LEGEND
FIRST PHASE 1956-7
SECOND PHASE 1957-8
THIRD PHASE 1958-61
ELEMENTARY SCHOOL
JR. HIGH SCHOOL
SENIOR HIGH SCHOOL
APPROX SCALE IN MILES
0 1 2 3 4 5
(A)
(B)

Fig. 1. Washington County television network

with an outer jacket of polyethylene. The Spirafil coaxial cable consists of a 9-gauge solid copper center conductor, a solid polyethylene filament helix and a 1/2-inch outside diameter aluminum conductor.

Both types of cables are spliced and terminated in special fittings utilizing special tools recommended by the manufacturer. The through splice and terminal fittings are shown in Figs. 8 and 9. Both types of cables are lashed to *6M* strand in the aerial portions of the network; only the JEL 100 cable is placed in underground conduits. The splices are enclosed in a *1A* splice case (Western Electric) normally used for polyethylene covered cables.

R-F Line Amplifier

A broad-band 75-ohm r-f amplifier having a band width of from 15 to 108 mc and a maximum gain of 33 (decibels) (db) at 88 mc is utilized in this system. The amplifier,[1] (Model 211, Spencer-Kennedy Laboratories), has gain-frequency characteristic which includes a positive slope of approximately 3 db between 54 and 88 mc. The amplifier has a noise figure of approximately 9 db at 88 mc. Its gain is controlled over 16 db range by either manual gain control or by an external bias supplied from an automatic level control panel. The unit is a chain amplifier utilizing 12 *6AK5*, or the ruggedized type number 5654, or the Western Electric *403B* vacuum tubes.

Automatic Level Control

A pilot generator installed at the Board of Education transmits a 77-mc pilot signal around the entire system. This actuates an automatic level control panel at every third amplifier point to automatically adjust the signal level on the system. At the level control points, an auxiliary amplifier is bridged to the through line by means of a directive coupling device called a "Chromatop" or a dividing network. The output of the auxiliary amplifier is coupled through a pilot band-pass filter to the automatic level control panel. The automatic level control panel rectifies the pilot signal and produces a bias for regulating the gain of the through line amplifier which compensates for the level changes in the three preceding sections.

R-F Filters

Two types of r-f filters are utilized in this system, one for band eliminating and one for band passing. The band eliminating filters (BEF) are used to block the television channel at all points of origination so as to effectively open the round robin thus preventing singing and "ghosts" from appearing in the television picture. These units are switched in and out by means of coaxial relays. A BEF for the pilot is permanently connected in the round robin at the Board of Education to block the pilot from repetitive trips around the system. Other BEF filters are utilized at the Board of Education location and at the Library to block the unused portions of the r-f spectrum, thus preventing singing and noise build-up at these frequencies. The channel 6 band elimination filter is designed to block channel 6, as well as the unused portion of the system's band width (88 to 108 mc). A band-pass filter for the pilot frequency is utilized at each of the automatic level

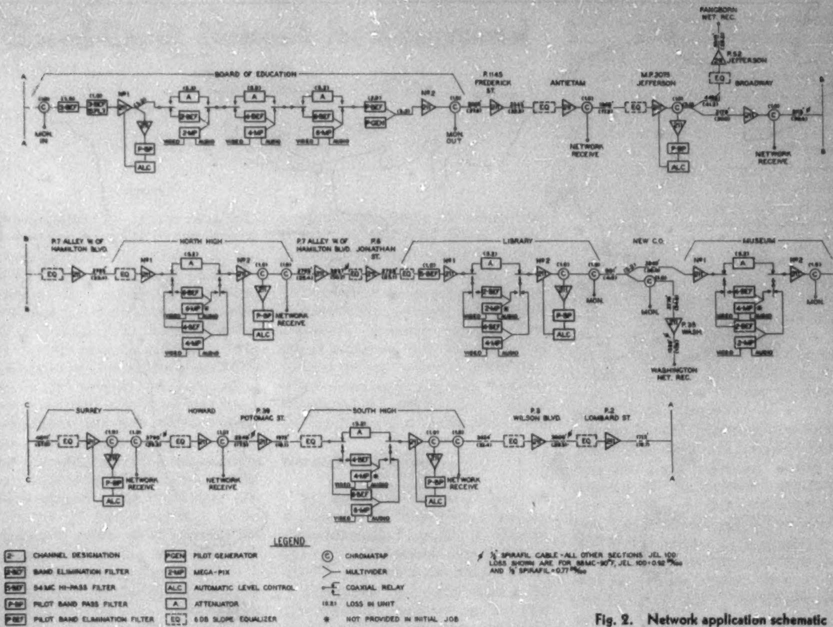

A
BOARD OF EDUCATION
MON. IN
VIDEO
AUDIO
MON. OUT
P.1145 FREDERICK ST.
ANTIETAM
NETWORK RECEIVE
M.P.3075 JEFFERSON
PANGBORN NET. REC.
P.52 JEFFERSON
BROADWAY
NETWORK RECEIVE
B
P.7 ALLEY W. OF HAMILTON BLVD.
NORTH HIGH
P.7 ALLEY W. OF HAMILTON BLVD.
P.8 JONATHAN ST.
NETWORK RECEIVE
LIBRARY
NEW C.O.
MON.
P.35 WASH.
WASHINGTON NET. REC.
MUSEUM
MON.
C
SURREY
NETWORK RECEIVE
HOWARD
NETWORK RECEIVE
P.39 POTOMAC ST.
SOUTH HIGH
NETWORK RECEIVE
P.3 WILSON BLVD.
P.2 LOMBARD ST.
LEGEND
2 CHANNEL DESIGNATION
2-BEF BAND ELIMINATION FILTER
5-BEF 54 MC HI-PASS FILTER
P-BP PILOT BAND PASS FILTER
P-BEF PILOT BAND ELIMINATION FILTER
PGEN PILOT GENERATOR
2-MP MEGA-PIX
ALC AUTOMATIC LEVEL CONTROL
A ATTENUATOR
EQ 6 DB SLOPE EQUALIZER
© CHROMATAP
MULTIVIDER
COAXIAL RELAY
(8.8) LOSS IN UNIT
* NOT PROVIDED IN INITIAL JOB
½" SPIRAFIL CABLE - ALL OTHER SECTIONS JEL 100
LOSS SHOWN ARE FOR 88MC-90°F, JEL 100=0.92 db/100
AND ½" SPIRAFIL=0.77 db/100

Fig. 2. Network application schematic

control points to tap off the pilot signal for the automatic level control operation.

Transmitting Terminals

The transmitting terminal includes a unit known as a MEGA-PIX, (Kay Electric Company) which converts the video and audio signals up to the r-f frequency channel assignment. The units are of two types, namely one that transmits on a single television channel and another that is tunable to any of the 12 television channels. The multichannel unit is mounted in a cabinet and is used for testing and maintenance.

Miscellaneous

Triaxial patch cords were utilized throughout the entire system to prevent r-f radiation or interference to the system. The connectors associated with these patch cords are of a wide variety to match the various connectors provided by the manufacturers. It is believed that the requirements for any future system should include standardized connectors.

Dividing networks with two branches are employed throughout the system. Each branch has a nominal impedance of 72 ohms and an insertion loss of 3.2 db for each branch and an isolation of 13 db between any two legs. These units are called Multividers, Spencer-Kennedy Laboratories). A directional coupler called a Chromatap (Spencer-Kennedy) is used to tap signals from the r-f system for school feeds, monitoring, and pilot regulating purposes. These couplers have approximately 1 db insertion loss to the line and a loss of 10 db to the tap.

Fixed 6-db slope equalizers are utilized,

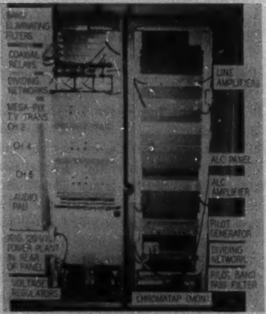

BAND ELIMINATING FILTERS
COAXIAL RELAYS
DIVIDING NETWORKS
MEGA-PIX T.V. TRANS CH 2
CH 4
CH 6
AUDIO PAD
90(120 V) POWER PLANT IN REAR OF PANEL
VOLTAGE REGULATORS
LINE AMPLIFIER
ALC PANEL
ALC AMPLIFIER
PILOT GENERATOR
DIVIDING NETWORK
PILOT BAND PASS FILTER
CHROMATAP (MON)

Fig. 3 (left). Multichannel program originating and line amplifier bays

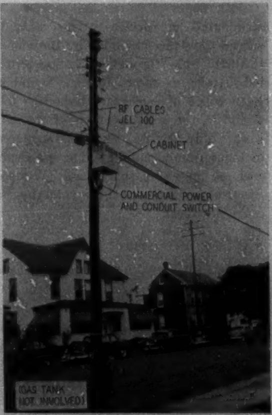

RF CABLES JEL 100
CABINET
COMMERCIAL POWER AND CONDUIT SWITCH
GAS TANK NOT INVOLVED

Fig. 4 (right). Single outdoor line amplifier installation

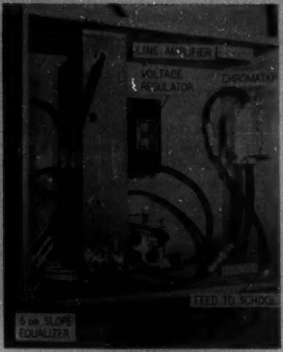

LINE AMPLIFIER
VOLTAGE REGULATOR
CHROMATAP
FEED TO SCHOOL
6 db SLOPE EQUALIZER

Fig. 5. Single indoor line amplifier

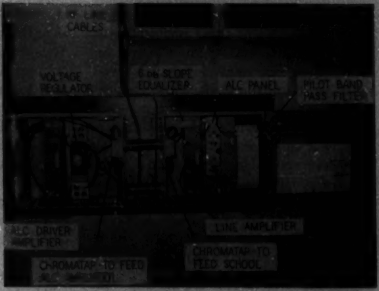

CABLES
VOLTAGE REGULATOR
6 db SLOPE EQUALIZER
ALC PANEL
PILOT BAND PASS FILTER
ALC DRIVER AMPLIFIER
LINE AMPLIFIER
CHROMATAP TO FEED SCHOOL
CHROMATAP TO FEED

Fig. 7. Automatic level control installation

LINE SPLICE CONNECTORS
TERMINATING CONNECTORS
CONNECTOR FOR TRIAXIAL PATCH CORD TO AMPLIFIER

Fig. 8. JEL 100 triaxial cable with line and terminal connectors

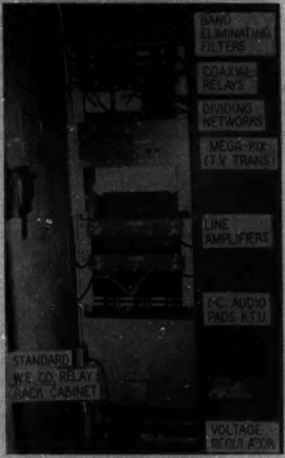

BAND ELIMINATING FILTERS
COAXIAL RELAYS
DIVIDING NETWORKS
MEGA PIX (T.V. TRANS)
LINE AMPLIFIERS
I-C AUDIO PADS KTU
STANDARD W.E. CO. RELAY RACK CABINET
VOLTAGE REGULATOR

Fig. 6 (left). Single channel program originating and line amplifier bay

LINE SPLICE CONNECTORS
TERMINATING CONNECTORS
TYPE "N" CONNECTOR

Fig. 9 (right). 1/2-inch Spirafil coaxial cable with line and terminal connectors

as determined by tests, to adjust the r-f levels in the band between 54 and 88 mc. These equalizers will be replaced or supplemented with units to cover the entire band of the system 15 to 88 mc to meet the future six television channel requirements.

Switching functions are provided by type *CR25N* coaxial relays (Danbury-Knudsen, Incorporated). The units are of 50-ohm impedance, equipped with BNC connectors and have 24-volt d-c windings. The relays are desired to ground the open contacts providing a 70-db crosstalk coupling path to the unwanted signal.

Test Equipment

The test equipment includes two types of sweep generators, two r-f field strength meters (Jerrold), a portable television receiver, a television tuner (Conrac) with video and audio output, a television monitor, oscilloscope, tube tester, and volt-ohmmeters.

Power

All units except the coaxial relays operate on 117-volt 60-cycle power. Voltage regulators (Raytheon) of 250 watts, 120 watts, and 60 watts are utilized at all installations to secure maximum stability for the system. All power feeds for the various amplifiers and input equipments are separately fused and provided with power switches. Three wire convenience outlets are provided at all locations for the operation of test equipment, lights, and soldering iron.

Performance

The system's performance has been generally satisfactory and during the first 7 months of operation there have been only two out-of-service failures, both of a mechanical nature.

The over-all qualitative evaluation of the system is determined by comparing the quality of a picture as viewed on two standard television sets, located side by side, one connected to the input and one to the output of the round robin. Resolution of the picture does not seem to be impaired by the transmission over $10^1/_4$ miles of cable and 21 amplifiers in tandem. There is, however, a discernible but not objectionable amount of noise in the picture. The quality of the sound channel as determined by the listening tests would be rated as from fair to good for voice and fair for music.

Plans for the expansion of the network for the second and third phases of the project include certain technical improvements that will result in an improved signal-to-noise ratio in the video channel as well as improvement in the sound channel. The signal-to-noise improvement will be obtained by increasing the radio frequency signal level from a minimum of zero decibels referred to one millivolt to a minimum of approximately +6 to +8 decibels referred to one millivolt. Improvement in the sound quality will be obtained principally in the transmitter terminal.

Plans for expansion also include the provision of three additional television channels and approximately 128 additional miles of network as shown in Fig. 1. The three additional channels will be originated on the regular assigned television channel frequencies of channels 7, 9, and 11. These channels will then be translated downward, at the studios, to frequencies in the spectrum between 15 and 54 mc for transmission over the network and then retranslated upward to their assigned frequencies at the schools.

Alarm and order wire facilities will also be included as a part of the network expansion. Low-signal alarms originating at each automatic level control amplifier will be transmitted over regular telephone facilities to a central point to indicate the location of abnormal conditions on the network. A party line order wire circuit will interconnect all amplifier and control locations to provide communication facilities during network alignment and maintenance work.

Results in the first stage of this experiment indicate that a radio frequency distribution system provides an adequate medium for the transmission of multichannel video and audio channels for short-haul, closed-circuit operation.

Reference

1. 200-MC TRAVELING-WAVE CHAIN AMPLIFIER, H. G. Rudenberg, Fitzroy Kennedy. *Electronics*, New York, N. Y., vol. 22, Dec. 1949, pp. 106–09.

Measurement of Voltage Ratio at Audio Frequencies

WILBUR C. SZE
NONMEMBER AIEE

HIGH orders of accuracy in voltage-ratio measurements are becoming increasingly important in computer fields. Accuracies of 0.01% or better are desired. Equipment such as precision ratio transformers are used in these applications. However, methods have been lacking for verifying the ratios of this equipment to that accuracy at the desired frequencies of 100 to 10,000 cps (cycles per second).

The purposes of this paper are: 1. to describe the use of a resistance-capacitance voltage divider in the measurement of voltage ratio and phase angle at audio frequencies, 2. to describe the special equipment used for the construction of the circuit, 3. to analyze various factors affecting the accuracy with which precision measurements can be made, and 4. to show the test results obtained from 100 to 10,000 cps for a ratio transformer.

Description of Method

The measuring circuit, shown in Fig. 1(A), consists of a symmetrically arranged resistance-capacitance network ($R_1C_1R_2$ and C_2) and the guard circuit consisting of a similarly arranged network ($R_{g1}C_{g1}R_{g2}$ and C_{g2}). The measuring circuit is balanced by adjusting either R_1C_1 or R_2C_2 until a null is indicated on the detector D. Similarly the guard circuit is balanced by means of adjusting $R_{g1}C_{g1}$ or $R_{g2}C_{g2}$. These procedures of balancing are continued successively until the detector D indicates a null in both the test and the guard circuit.

In addition to these two circuits there are two auxiliary circuits: 1. the Schering bridge, Fig. 1(B), used for the precise measurement of the phase angles and 2. the Wheatstone bridge, Fig. 1(C), used for the precise determination of the voltage ratios.

The transfer from one circuit to another is accomplished by the arrangement of switches shown in Fig. 2, described later.

The guard circuit, for both the transformer test circuit and the Schering bridge, is arranged to eliminate the effect of ground admittances by completely shielding the detector, detector circuit, and associated apparatus.

Theoretical Relations

The equations for the ratio-correction factor and phase angle were developed as follows: By definition, the voltage ratio of a transformer or divider is the ratio of the rms primary or input terminal voltage $|V_p|$ to the rms secondary or output terminal voltage $|V_s|$ so that

$$\frac{|V_p|}{|V_s|}=NF$$

where N is the nominal ratio and F is the ratio-correction factor derived from the measured quantities. The phase angle γ is the difference in phase between the primary and the secondary voltage of the transformer or divider.

The circuit diagram for the derivation of the following equations is shown in Fig. 1(A).

$$Z_1=\frac{R_1}{1+j\omega C_1R_1}$$

$$Z_2=\frac{R_2}{1+j\omega C_2R_2}$$

When detector D indicates a null, then $I_s=0$, $V_p=I_p(Z_1+Z_2)$, and $V_s=I_pZ_1$. Let

$$V_p=|V_p|(\cos\gamma-j\sin\gamma)$$

$$V_s=|V_s|=|V_p|/NF$$

Then

$$NF(\cos\gamma-j\sin\gamma)=\frac{V_p}{V_s}=\frac{Z_1+Z_2}{Z_1}=\frac{R_1+R_2+j\omega R_1R_2(C_1+C_2)}{R_1(1+j\omega R_2C_2)} \qquad (1)$$

From equation 1 equate the real terms to get

$$F=\frac{1}{N}\left(\frac{R_2}{R_1}+1\right)\left(\frac{1}{\cos\gamma+\omega R_2C_2\sin\gamma}\right)$$

Since the a-c resistors R_1 and R_2 are not stable to the degree of accuracy desired, the ratio of the divider R_2/R_1 is deter-

---

Paper 57-648, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 22, 1957; made available for printing April 17, 1957.

WILBUR C. SZE is with the National Bureau of Standards, Washington, D. C.

The author wishes to acknowledge the valuable help and advice received from F. L. Hermach of the National Bureau of Standards.

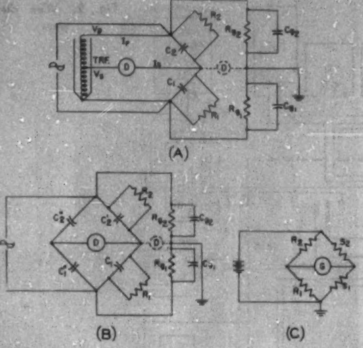


Fig. 1. Circuit diagrams

A—Transformer calibration circuit
B—Schering bridge circuit
C—Wheatstone bridge circuit

mined immediately after each measurement by using the highly accurate d-c resistance bridge shown in Fig. 1(C). Then

$$F=\frac{1}{N}\left(\frac{S_2}{S_1}+1\right)\left(\frac{1}{\cos\gamma+\omega R_2C_2\sin\gamma}\right)$$

However, in order to obtain the desired accuracy, a-c/d-c corrections for R_2 and R_1 must also be added. The previous equation therefore needs to be modified as follows

$$F=\frac{1}{N}\left[\frac{S_2}{S_1}(1+\mu_{R2}-\mu_{R1})+1\right]\times \left(\frac{1}{\cos\gamma+\omega R_2C_2\sin\gamma}\right) \quad (2)$$

where μ_{R2} and μ_{R1} are the a-c/d-c corrections for resistors R_2 and R_1 respectively. Although these corrections are very small they cannot be neglected. They arise from skin effect, dielectric loss, and the residual reactances which are inevitably associated with all resistors. These effects become increasingly important as the test frequency is raised, because they change the terminal resistance values from their low-frequency or d-c values. Measurements to determine μ_{R1} and μ_{R2} are described in a later section.

Again from equation 1 equate the imaginary terms to get

$$jNF(\omega R_2C_2\cos\gamma-\sin\gamma)=j\omega R_2(C_1+C_2)$$

Since γ is usually very small, it can be assumed that in this expression

$$NF\approx(R_1+R_2)/R_1,\ \cos\gamma\approx1,\ \sin\gamma=\gamma$$

Therefore

$$\gamma=\frac{\omega R_2}{R_1+R_2}(C_2R_2-C_1R_1)\text{ radians} \quad (3)$$

Because it is nearly impossible to measure accurately the stray capacitances of the test circuit which are in parallel with the capacitors C_1 and C_2, errors are introduced in the determination of the phase angle γ if equation 3 is used. A much more elaborate substitution method was therefore developed to eliminate or minimize the aforementioned sources of error so that they do not affect the desired accuracy of the results. For the connection shown in Fig. 1(A), a null balance is determined. The values of R_1, R_2, and C_2 are then left fixed, and C_1'' and C_2'' are connected in place of the transformer, as shown in Fig. 1(B). This is the well-known Schering bridge circuit with the special requirement that the phase defect angles of capacitors C_1'' and C_2'' must be negligible. These must be very stable capacitors but their values of capacitance need not be known with great accuracy. Three-terminal air capacitors having phase defect angles of less than 2 seconds (and therefore negligible) were constructed as will be described in a later section.

A new null balance of the bridge is obtained by varying C_1'' or C_2'' and by making a slight adjustment of C_1 to a new value C_1'. Then

$$R_2/R_1=C_1'/C_2=C_1''/C_2''$$

In this particular physical arrangement, the values of the capacitors C_1 and C_2 and of resistors R_1 and R_2 remain essentially fixed with respect to each other except for a very small adjustment in C_1.

Because the phase angle γ will be determined as a slight difference in setting from C_1 to C_1', the exact values of C_1 and C_2 may remain undetermined as long as these capacitors are stable and do not shift in value. Similarly the stray capacitances of the circuit may remain undetermined as long as they remain fixed and may therefore be considered as included in the settings of C_1 and C_2.

Substitute $R_2=R_1(C_1'/C_2)$ in equation 3 to obtain

$$\gamma=\frac{R_1R_2}{R_1+R_2}\omega(C_1'-C_1)\text{ radians} \quad (4)$$

By the same token, C_2 may be adjusted to C_2' rather than varying C_1. The following is then obtained

$$\gamma=\frac{R_2^2}{R_1+R_2}\omega(C_2-C_2')\text{ radians} \quad (5)$$

Shielding and Grounding

The use of the Wagner ground is an easy and simple way of eliminating the effect of ground admittances. Consider the impedance bridges of Figs. 1(A) and (B), consisting of the usual 4-arm network, but having in addition the auxiliary arms Z_{g1} and Z_{g2} with a ground at their junction. The latter forms the auxiliary impedance bridge with the detector circuit shown by dashed lines.

Although the entire detector circuit is at ground potential when both test and guard circuits are balanced and all ground-admittance effects have been eliminated, a potential difference exists between the detector circuit and the leads from the supply. To eliminate inductive pick-up loops, these supply leads must be placed as close as possible to the detector circuit; thus the effect of leakage between them may be considerable. Therefore, the entire detector circuit must be shielded (shields shown by dashed lines in Fig. 2) to eliminate or reduce this effect to such an extent that it does not affect the accuracy of the results.

Wiring Diagram

Fig. 2 shows the actual wiring diagram of the setup and the arrangement of the switches in the transfer switch box. It is important to note that this transfer switch box is separated into two shielded compartments. The smaller compartment contains the entire detector circuit which, when properly balanced, is at ground potential. The larger compartment contains the a-c/d-c power selector switch and Schering bridge, transformer test selector switches.

Every precaution must be taken to ensure that: 1. components of the entire circuit be arranged as symmetrically as possible, 2. all leads in the test circuit be kept as short as possible, 3. the connecting leads, between the switch box and any group of apparatus, be tightly twisted together, 4. all the shields be electrically connected, and 5. that there be only a single ground point in the whole circuit, i.e., the Wagner ground.

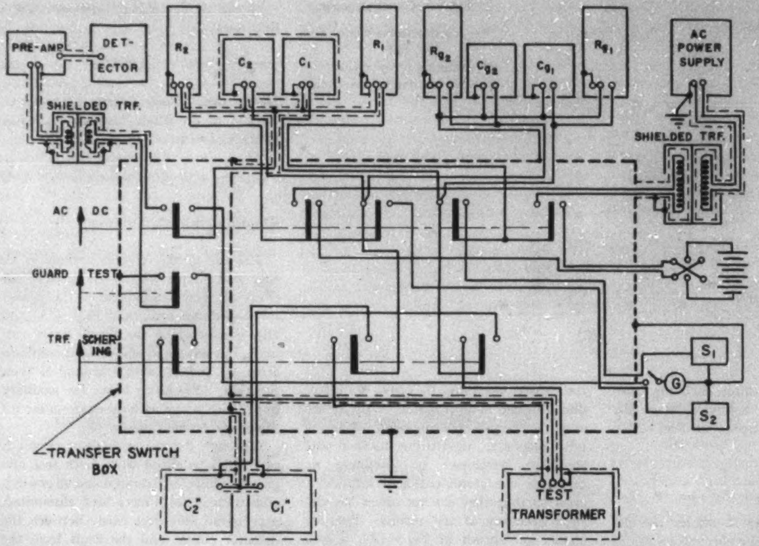


Fig. 2. Wire diagram

Test Procedure

The procedure in testing a voltage transformer or voltage divider with this circuit arrangement consists of three separate steps.

Step 1. Connect the test transformer in the circuit and set the selector switches in Fig. 2 to marks *AC*, *Test*, and *Trf*, essentially giving the circuit shown in Fig. 1(A). Next select the proper values of settings for all the apparatus in the circuit, so that

$$S_2/S_1 \approx R_2/R_1 \approx R_{g2}/R_{g1} \approx C_1/C_2 \approx C_1''/C_2'' \approx C_{g1}/C_{g2} \approx N-1$$

Vary either C_1 and R_1 or C_2 and R_2 to obtain a null indication on the detector. Now set the center switch to *Guard* and adjust C_{g1} and R_{g1} or C_{g2} and R_{g2} to obtain a null indication on the detector. This procedure must be repeated until no further adjustment is required for a balance indication of the detector with the switch set to either the *Test* position or the *Guard* position.

Step 2. Turn the top selector switch in Fig. 2 to *DC;* this essentially gives the same circuit as shown in Fig. 1(C). Close the galvanometer switch and vary S_1 or S_2 for a zero-current indication of the galvanometer *G*, with the battery connections reversed.

Step 3. Reset the selector switches to *AC*, *Test*, and *Schering*. This gives the circuit of Fig. 1(B). Obtain a null indication on the detector by varying C_2'' and C_2 or C_1'' and C_1. Again the guard circuit must be balanced as in step 1.

Steps 1, 2, and 3 should follow one another with as little delay as possible between each step, so that the variation in resistance values of R_1 and R_2 due to temperature change is small.

With the values of settings recorded from the procedure, the voltage-ratio correction factor and the phase angle of the test transformer can be computed by substituting the proper values into equations 2 and 4 or 5, depending upon whether C_2 or C_1 was varied.

Description of Apparatus

The two low-loss variable air capacitors C_1'' and C_2'' are *NBS*-type, parallel-plate, 100 to 5,000 picofarads (micromicrofarads), modified for use as shielded 3-terminal capacitors. Their general construction is shown in Fig. 3. One set of the plates is mounted on the central shaft and may be rotated by the shaft relative to the fixed set. The central shaft is electrically connected to the case. The other set is stationary and supported on three guarded quartz-rod insulators. The guards are arranged to shield completely the quartz insulation from the active field between the working electrodes. One of these supporting structures is shown in Fig. 4. The dielectric loss angle of these capacitors was measured and found to be about 1 second at 1,000 cps.

C_1 and C_2 are two precision variable air capacitors specially developed and designed for this setup, 20 to 120 picofarads and 50 to 1,100 picofarads respectively, used as shielded 3-terminal capacitors. The dielectric loss angles were measured and found to be approximately 2 seconds and 5 seconds respectively at 1,000 cps. Such small values are relatively unimportant in balances of the circuit as long as the corresponding conductance values are very small in comparison with the resistors R_1 and R_2.

Resistors R_1 and R_2 are two precision a-c decade resistors, 11,111 ohms total in steps of 0.1 ohm. Their resistance elements have very small residual reactances. However, in precise a-c measurements, especially at higher audio frequencies, the effect of these reactances should be taken into account. A special bridge described later was therefore set up to evaluate the change of the effective resistance over the range from direct current to 10,000 cps.

The null detector is connected to the several bridge arrangements through a preamplifier and an impedance-matching

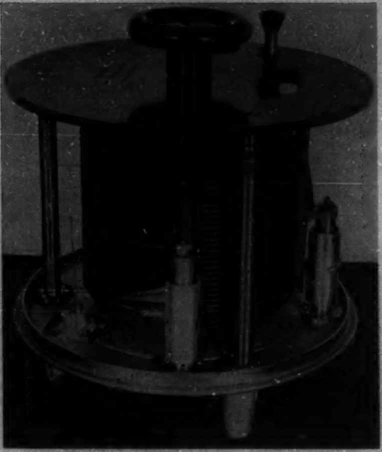


Fig. 3 (left). Three-terminal air capacitor, case removed

Fig. 4 (right). Shielded support for air capacitor

shielded transformer; see Fig. 2. The battery-operated preamplifier is a single pentode amplifier. It is completely shielded by its metal container connected to the shields for the detector leads. The impedance-matching shielded transformer had individually shielded windings.

A commercial wave analyzer was used in balancing the test and guard circuits. The over-all resolution of the complete detecting system is better than 1 microvolt.

D-C Wheatstone Bridge

Only those special features of the d-c Wheatstone bridge that are essential to a high order of accuracy of measurement are described.

Precision standard resistors were used for the S_1 and S_2 arms of the Wheatstone bridge. The resistance elements are hermetically sealed in oil to eliminate effects from the humidity variations in the laboratory.

The arrangement of standard resistors in the ratio arms S_2/S_1 are determined by the ratio R_2/R_1. These standard resistors are connected in series-parallel arrangements so that

$$_2/S_1 \approx N-1$$

An adjustable resistance box which is not hermetically sealed is connected in one arm, preferably in the arm with large resistance values, for the fine adjustments of the d-c bridge. The lowest dial increment on the resistance box must give sufficient adjustment for the accuracy of the results desired. Corrections for exact value, temperature coefficient, and lead resistances must be applied to all the resistors. Speical precautions must be taken to ensure that the leakage currents across the ratio arms are insignificant. With this procedure, the uncertainty of ratio S_2/S_1 should not exceed 20 ppm (parts per million) or 0.002%.

Changes in value of these standard resistors, since their initial measurements, have been trivial. As a guard against any significant unpredicted changes in resistance, a schedule has been established whereby these resistors and the resistance box can be intercompared to 10 ppm after about 2 hours of systematic work.

A-C Tests of Resistors

Since the voltage ratio is derived from measurements with d-c resistance standards, it is necessary to know the change of effective resistance of the a-c resistors with frequency. In this connection Berberich developed an excellent substitution method for determining the phase angle of a resistor.[1] By means of a bridge circuit he compares a resistor with a parallel-wire reference standard of computable reactance which has approximately the same resistance value as that of the resistor. The substitution method he developed makes it simple to minimize the effect of parasitic influences in the circuit, does not require special apparatus, is easy to use, and gives highly accurate results. It is readily adaptable for measuring the effective resistance of a resistor with change in frequency at the same time the phase angle is measured, without altering the test setup.

Fig. 5 shows a diagram of the circuit used for measuring the phase angle and effective resistance of the resistors R_1 and R_2. Capacitors C_{W2} and C_{W1} and a selector switch were added to the Wagner arms in order to provide for balancing the circuits at frequencies up to 10 kc. In this figure R_s' and L_s' represent the resistance and equivalent inductance of the parallel-wire standard. R_x' and L_x' represent the resistance and equivalent inductance of the resistance boxes whose phase angles are to be measured.

To determine the a-c/d-c difference of the resistance boxes the ohmic value of the unknown resistance box is set equal to that of the parallel-wire standard at a low base frequency, 400 cps or less, then without changing the original settings of the box, the frequency is increased in successive steps and the bridge circuit is rebalanced with arm C for both the standard and the box at each step.

The proportional change of the effective resistance of the computable parallel-wire standard α is computed from

$$\alpha = (2\omega^2 LC/3) - (2\omega^2 C^2 R^2/15)$$

The change β of the effective resistances of the resistance box due to frequency is computed from the difference of the resistance settings of R_x and R_s dials at sucessive increasing frequency, i.e.,

$$\beta = (R_x - R_s)/R_s$$

However, the effective resistance and

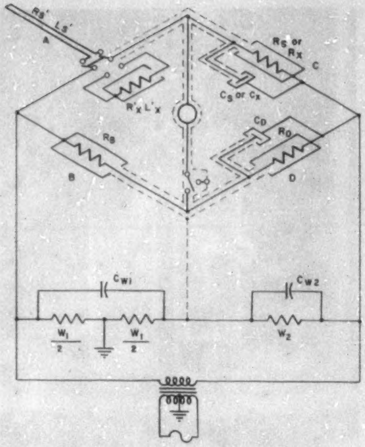


Fig. 5 (left). Modified Berberich's bridge

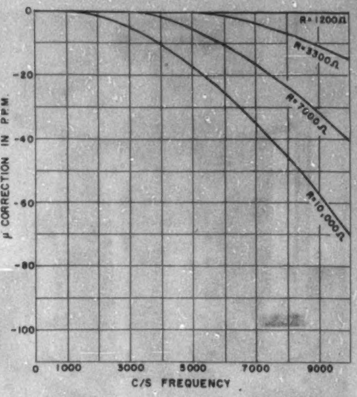


Fig. 6 (right). Frequency correction for GR-type 1432-N decade resistor

effective reactance of the box, $R'+jx'$, as determined by Berberich's method, are based on equivalent series circuit. When the box is actually connected and used in the ratio-measuring circuits (Figs. 1(A) and (B)), the effective parallel resistance and reactance enter into the formulas. Therefore, to the foregoing correction β must be added the term

$$\lambda^2=\omega^2 C_{par}{}^2 R_{par}{}^2$$

where C_{par} is the measured equivalent parallel capacitance and R_{par} is the equivalent parallel resistance; see the Appendix.

The effective resistance of the resistance box at any frequency may be expressed in the following equation

$$R_x''=R_{x1}'(1+\alpha+\beta+\lambda^2)=R_{x1}'(1+\mu)$$

The correction μ is expressed in proportional parts. R_{x1}' is the resistance value of the resistance box at base frequency. u is shown as μ_{R1} for corrections R_1 and μ_{R1} for correction to R_2 in equation 2, corresponding to measured values that were read from Fig. 6.

Results of Tests

Test results of one ratio transformer at an average ambient temperature of 26 degrees centigrade and with 50 volts rms applied to the primary are given in Table I. The transformer has input-to-output ratio settings covering the range from 0.000000 to1.000000, adjustable by means of 5-decade switches and one rheostat.

Conclusions

A setup has been developed for measuring the ratio and phase angle of precision ratio transformers or alternating voltage dividers at audio frequencies. An overall accuracy of within about 0.005% in ratio and 0.2 minute in phase angle at frequencies from 100 to 10,000 cps is achieved. To obtain this accuracy in ratio, corrections must be applied for the a-c and d-c difference of the decade resistance boxes and for the differences between true and nominal values of ohmic resistance of the standard resistors. The accuracy of phase-angle measurements relies on the constancy of the small loss angle of the precision air capacitors, which should not change appreciably except at high relative humidities and because of exceptional surface or air contamination. Ordinary care is sufficient to avoid bothersome changes.

A simple circuit is arranged symmetrically with the detector always at earth potential. Identical apparatus is provided on either side the detector circuit so that it can be interchanged electrically. In order to ensure that the circuit is electrically as well as physically symmetrical, the test transformer or divider

Table I. Results of Test

Nominal Ratio	100 CPS		400 CPS		1,000 CPS		3,000 CPS		7,500 CPS		10 Kc	
	F	γ, Minute	F	γ, Minute	F	γ, Minute	F	γ, Minute	F	γ, Minute	F	γ, Minute
10/0.39					1.00004	−0.5						
10/0.5					1.00003	−0.4	1.00006	+0.3	1.00015	+0.8	1.00025	+1.1
10/0.5333					1.00004	−0.4						
10/0.5777					1.00003	−0.4						
10/0.7					1.00002	−0.3						
10/1	1.00000	+0.8	1.00001	−0.8	1.00002	−0.1	1.00004	+0.3	1.00013	+0.8	1.00022	+1.0
10/2					1.00001	0.0						
10/3					1.00001	−0.1						
10/4					1.00000	−0.1						
10/5	1.00000	+0.3	1.00000	−0.1	1.00000	0.0	1.00000	0.0	1.00001	−0.1	1.00002	+0.3
10/6					1.00000	−0.1						
10/7					1.00000	−0.1						
10/8					1.00000	−0.1						
10/9	1.00000	0.0	1.00000	0.0	1.00000	0.0	1.00000	−0.1	0.99999	−0.2	0.99999	−0.2

can also be inverted in relation to the rest of the circuit.

The equipment has been used in checking transformers and a-c dividers for voltage ratio up to 1,000 to 1. The flexibility of the setup is considerable. Fractional ratios may be measured as readily as integral ratios. Test results for high and low ratios are obtained with a slight reduction in the accuracy (±0.005%) attainable for ratios between 1 to 1 and 100 to 1.

Appendix. Derivation of Equations for Correction μ

Correction μ is the sum of α, β, and λ^2. α is the correction due to the change of effective resistance of the parallel-wire standard with change in test frequency. As derived by Harris, it is, to the desired approximation,[2]

$$\alpha=\frac{2\omega^2LC}{3}-\frac{2\omega^2C^2R^2}{15}$$

β is the correction due to the change of effective resistance of the resistance box by the change in test frequency. Consider Fig. 5, with the standard connected in arm A of the bridge, then at any frequency

$$R_s'=R_s\left(\frac{R_B}{R_D}\right)$$

With the same frequency, substitute a resistance box in arm A, and

$$R_x'=R_x\left(\frac{R_B}{R_D}\right)=\frac{R_xR_s'}{R_s}$$

since $R_x-R_s<<1$. Then

$$R_x'-R_{x1}'=\frac{R_xR_s'}{R_s}-\frac{R_{x1}R_{s1}'}{R_{s1}}$$

where subscript 1 indicates the base frequency. Since $R_{x1}=R_{s1}$ was set, and $R_s'=R_{s1}'(1+\alpha)$, then

$$R_x'-R_{x1}'=R_{s1}'\left(1+\frac{R_x-R_s}{R_s}\right)(1+\alpha)-R_{s1}'$$

Let

$$\beta=\frac{R_x-R_s}{R_s}$$

Then

$$R_x'\approx R_{x1}'+R_{s1}'(\alpha+\beta)$$

$$R_x'\approx R_{x1}'\left[1+\frac{R_{s1}'}{R_{x1}'}(\alpha+\beta)\right]$$

$$R_x'\approx R_{x1}'(1+\alpha+\beta)$$

since $R_{s1}'/R_{x1}'\approx 1$. R_x' is ohmic resistance expressed as in the equivalent series circuit.

The term λ^2 is the changeover factor from the equivalent series to equivalent parallel connections:

$$R_{ser}+jX_{ser}=\frac{R_{par}+j\dfrac{R_{par}^2}{X_{par}}}{1+\dfrac{R_{par}^2}{X_{par}^2}}$$

Let

$$R_{par}/X_{par}=\lambda\ll 1$$

Then

$$R_{ser}+jX_{ser}=\frac{R_{par}(1+j\lambda)}{1+\lambda^2}$$

from which $R_{par}=R_{ser}(1+\lambda^2)$ and $\lambda=X_{ser}/R_{ser}$. Therefore, the effective resistance for equivalent parallel circuit at any frequency is

$$R_x''=R_x'(1+\lambda^2)=R_{x1}'(1+\alpha+\beta)(1+\lambda^2)$$

$$R_x''\approx R_{x1}'(1+\alpha+\beta+\lambda^2)=R_{x1}'(1+\mu)$$

R_{x1}' is the resistance value at base frequency.

References

1. The Measurements of the Phase Angles of Shielded Resistors, L. J. Berberich. *Physics*, New York, N. Y., vol. 3, Dec. 1932, pp. 296–313.

2. Electrical Measurements (book), F. K. Harris. John Wiley & Sons, Inc., New York, N. Y., 1952, pp. 217–20.

3. Nature of Energy Losses in Air Capacitors at Low Frequencies, A. V. Astin. *Journal of Research*, National Bureau of Standards, Washington, D. C., vol. 22, 1939, pp. 673–95.

4. Design, Construction and Use of Resistors of Calculable Reactance, N. F. Astbury. *Journal*, Institution of Electrical Engineers, London, England, vol. 76, 1935, p. 389.

5. A.C. Bridge Methods (book), B. Hague. Isaac Pitman and Sons, Ltd., London, England, 1946.

6. Instrument Transformer (book), B. Hague. Isaac Pitman and Sons, Ltd, 1936.

An Electronic Electrometer as a Versatile Corona Detector

J. G. ANDERSON
MEMBER AIEE

J. S. KRESGE
ASSOCIATE MEMBER AIEE

Synopsis: A new technique is described for the detection and estimation of magnitudes of corona discharges in electric devices during high-voltage testing. This technique employs the electronic electrometer as an instrument to measure the residual charge left in high-voltage structures by both impulse and low-frequency corona or breakdown discharges. For impulse testing, the method can be used with any shape or duration of voltage wave, and for coil stacks it has permitted the prediction of permanent damage and breakdown in many cases. The method is applicable in the testing of high-voltage structures involving at least some solid insulation.

A MAJOR OBJECTIVE of the electrical design and development engineer is to obtain a clear and reliable picture of the behavior of electric apparatus during overvoltage testing. A factor of great importance which should be determined as accurately as possible is the onset level of corona in any particular device. An estimate of the relative corona magnitudes at various higher voltage levels is also of value in predicting the ultimate failure of the apparatus, and in estimating the degree of insulation damage caused by the corona discharges.

In the past, the basic circuit used for most discharge detection schemes has been the Quinn[1] circuit or a variation thereof. Corona currents in the test piece cause small current variations in the terminal leads, and these variations are amplified, shaped, and fed to an oscilloscope where the corona currents appear as pulses. Improvements in this technique by balancing out the test-piece displacement or conductance currents have been described by Hopkins, Walters, and Scoville,[2] Hagenguth and Liao,[3] and others. However, even these improved techniques are of little value for certain applications. For example:

1. Separation of the corona currents and normal charging currents becomes very difficult for high-capacitance test pieces and/or rapidly varying applied voltages.

2. External circuit corona or sparking, which is often impractical to eliminate completely, may produce the same type of

---

Paper 57-632, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted November 16, 1956; made available for printing April 12, 1957.

J. G. Anderson and J. S. Kresge are both with the General Electric Company, Pittsfield, Mass.

The authors acknowledge the support of this work by J. H. Hagenguth, A. F. Rohlfs, and T. R. Walters, and the many suggestions and contributions of T. W. Liao. It has also been because of the active support and enthusiasm of the authors' coworkers in the General Electric High Voltage Laboratory that this work has progressed thus far.

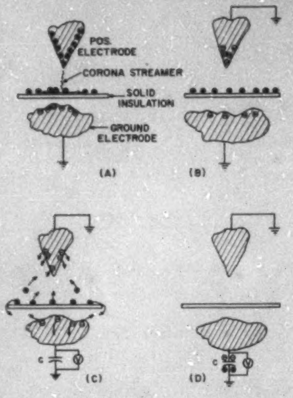


Fig. 1. Charge motion in a simple dielectric structure, and its detection with an electronic electrometer

corona detector response as the test-piece corona one wishes to measure.

3. The high-voltage impulse testing of coil stacks or transformers may require external circuits which conflict with the techniques required for corona detection.

4. The corona pulses within windings may be so severely attenuated by their passage through the winding capacitance network that their amplitudes at the ends of the winding are negligible.[4]

To overcome these and other difficulties, it was desirable to develop corona detection techniques which worked on the "memory" principle, in that remanent effects of the corona would be measured after the high-voltage testing surges interfering with normal corona measurements had been removed or decreased to a negligible value. This required instruments which measured only those corona effects persisting after the applied impulse wave had died away. A very useful technique of this type developed by one of the authors[5] was an ultrasonic method of corona detection and location. This method measured the ultrasound produced by electrical discharges and since this ultrasound requires many hundreds of microseconds to traverse a few feet it does not arrive at a transducer until after all high-voltage disturbances are over.

Another technique using the "memory" principle is the electrometer method described in this paper. This method is based on measurement of the residual charge left in a high-voltage structure by previous corona discharges. From this residual charge the corona onset voltage can be determined, and measurements of order of magnitudes of the original corona discharges are easily possible in many cases. The significant success of this method during high-voltage tests of coil stacks and other devices indicates that the method may have wide application to laboratory discharge measurements on many other kinds of electric apparatus.

Theory of Electrometer Corona Detection

When corona impinges on solid insulation, a residual charge is left on the insulation surface.[6] This is illustrated in Fig. 1(A) for a simple electrode configuration with solid insulation interposed. The residual charge induces an equal and opposite charge on near-by ground electrodes, and when the test is completed and voltage is removed from the high-voltage electrode, the charge induced on the electrodes will redistribute as shown in Fig. 1(B). There immediately follows a relatively long period during which the stored charge in the structure "leaks off" and recombines, Fig. 1(C). This period may vary from a few seconds to several hours for good insulation. If during this time a capacitor is inserted in the current lead of the ground electrode, some of the induced charges that would otherwise move along this conductor to ground collect on the capacitor plates as shown in Fig. 1(D), creating a voltage according to the relation $Q=cv$. An infinite impedance voltmeter V across this capacitor would then deflect according to the magnitude of the stored charge. Thus if one were interested only in knowing whether or not corona had occurred, the insertion of the capacitor C and a reading of the infinite impedance voltmeter would supply an immediate indication. As will be shown later, it also supplies a measurement of the magnitude of the corona discharge.

Another type of residual charge measurement that can be used is shown in Fig. 2. A probe (perhaps consisting of a small metal plate) is placed near a region previously charged by a corona streamer. Then the capacitance C_1, from charged surface to probe and the external capacitance C_2 from probe to ground form a simple capacitance voltage divider, and a voltage V appears across C_2 proportional to the charge. However, one is rarely at liberty to probe around the interior of most electric apparatus. Therefore, in most practical configurations, the only available probes are the electrodes already present. This is essentially what was done in Fig. 1 since the ground electrode was disconnected from ground and used as a probe.

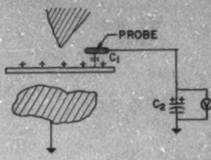


Fig. 2. Alternate method of detecting corona by using an additional probe

Neglecting the effects of dielectric absorption for the moment, the voltage across the electrometer capacitor C of Fig. 1(D) due to residual corona charge in the insulation structure can be computed. This is done in detail in Appendix I, and it is shown that the equivalent measuring circuit is similar to an impulse generator circuit with very long time constants. For most liquid-immersed insulations, the voltage across C reaches crest in a matter of many seconds or even minutes and then decays at a slower rate. This is a definite advantage, because it means that high-speed switching of C into the test-piece circuit is not necessarily required.

Appendix I shows that the maximum electrometer voltage and the initial rate-of-change of electrometer voltage are both directly proportional to the residual corona charge left in a simple test piece. It also shows that the electrometer voltage tends to rise to some maximum value and then decays slowly to zero. For more complicated test pieces, the simple relationships derived in Appendix I may not hold. However, the total corona charge is at least as great as that indicated by the maximum electrometer reading.

Instrumentation and Procedures

The most practical voltmeter for use in corona charge measurements is an electronic electrometer. This is a small vacuum tube d-c voltmeter having an input impedance of 10^{13} to 10^{14} ohms, depending on ambient humidity conditions. It is essential that such a high-impedance meter be used to prevent the discharge of the measuring capacitor as fast as the charge arrives. An electrometer with a wide range of voltage scales from a few millivolts to several volts is most suitable. The capacitor in parallel with the electrometer can also be varied to adjust the sensitivity, since for a given corona charge the capacitor voltage read by the electrometer is inversely proportional to the capacitance. An air capacitor (satis-

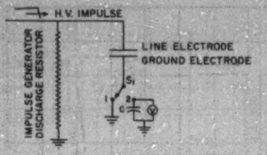


Fig. 3. Basic circuit for detecting impulse corona in capacitive test samples

factorily shielded from stray fields) is the best choice, but in many applications an ordinary mica capacitor is sufficient. Whatever type of capacitor is used, it must be reasonably free of dielectric absorption effects and must have a d-c leakage resistance approaching the 10^{13} to 10^{14} ohms of the electrometer. Precautions should be taken to minimize the generation of spurious circuit voltages by electrochemical action, stray fields, or other means. For all tests the electrometer and its capacitor *C* should be removed from the testing circuit, short-circuited, and grounded during the time high voltage is applied to the test piece. This is because the electrometer is easily damaged by overvoltages that might occur by premature failure of the device being tested. It is only after the high voltage has been removed that the electrometer reading is made.

Since it is stored charge that is measured after the test voltage is removed, the electrometer method of corona measurement can be used with equal ease for steep-front, chopped-wave, or full-wave impulse tests; or for any other unidirectional type of voltage application where the charge left in the test structure is considerably greater than the charge left by ordinary dielectric absorption. For most high-voltage devices, the dielectric absorption characteristics can be established by experience, and any sudden abnormal deviation from this experience at any given voltage level is an indication that corona charges are also being produced.

For low-frequency testing, or for the application of oscillatory impulse waves or switching surges, the electrometer method is also promising. Here, however, if both positive and negative corona discharges are present, one polarity discharge may partially neutralize previously deposited charges of the opposite polarity. Therefore, while in such cases the electrometer may show whether corona was present, it does not always indicate order of magnitudes of the discharges. For this reason, the method is best suited for short-duration unidirectional voltage tests where it indicates both the presence and magnitude of the corona discharges.

There are some applications for which the electrometer is not a practical discharge detector. For example, it is necessary that the insulation structure be such that residual charges will persist long enough to permit a reading. This destroys the effectiveness of this technique in the measurement of gaseous discharges between metallic electrodes where no solid insulation is present. The electrometer method also seems to be insensitive to many types of void corona, since recombination may occur within the void without induced charges being produced on the near-by-electrode surfaces. More work needs to be done in this area. As in all devices, this method has both advantages and limitations, and one can hardly expect any one device to solve all discharge detection problems that may arise.

Applications to Practical Test Samples

When devising the circuitry for detecting corona with an electrometer, due consideration must be given each individual test setup to determine which electrometer connection will give the best indication of corona in the sample. Then it is often necessary to apply physical reasoning to the readings obtained to determine what part of them are due to sample corona and what are due to other influences such as dielectric absorption or corona on associated equipment such as the entrance bushing. It is for these reasons that this technique is best suited for the developmental type of testing encountered in a laboratory and is not well adapted to routine factory-type tests. While it would be impossible to discuss the circuitry for all possible applications, a few examples will be discussed to illustrate the principles involved.

CAPACITANCE-TYPE TEST SAMPLES

For full-wave impulse tests on samples of relatively low capacitance (bushings, basic insulation structures for developmental tests, small coupling capacitors, etc.) the balanced impulse bridge technique discussed by Hagenguth and Liao[3] gives excellent results. However this method has not given satisfactory results during steep-front or chopped-wave impulse tests. The electrometer method of detection, since it works as a memory device, is applicable for all types of impulse waves applied and therein lies one of its great advantages. The circuit for use with capacitive test samples is illustrated in Fig. 3. The switch S_1 is arranged to connect the normally grounded test sample electrode to the electrometer circuit after application of the impulse wave. The line electrode, during the time the electrometer is in the circuit, is effectively grounded through the impulse generator discharge resistor. The principles of this circuit have been illustrated in Fig. 1.

The results obtained with this circuit during impulse tests on two high-voltage test samples are shown on Fig. 4. The samples were oil-immersed parallel-plane electrodes separated by solid insulation which extended about 12 inches beyond the electrode edges to provide a rather long creepage path. Steep-front, chopped-wave, and full-wave impulses were applied sequentially in the order given. For each level, the steep-front and chopped-wave crest voltages were respectively 1.5 and 1.15 times the full-wave crest voltage. Test levels were increased in approximately 5% steps until breakdown occurred. Electrometer readings taken after the application of each wave at each level are plotted on Fig. 4. The onset of corona is indicated when the electrometer reading suddenly increases to a much higher value. The relative corona intensity was considerably greater for the full wave than for the chopped wave or steep-front-type wave at any particular test level. This is a result which should be expected since the corona occurred as creepage streamers off the electrode edges. The physical distance of propagation and hence the resultant

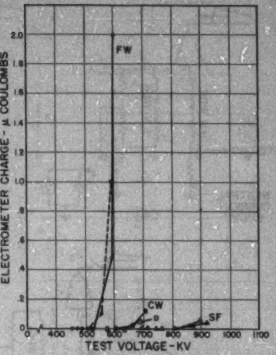


Fig. 4. Electrometer readings obtained during impulse tests on parallel-plane electrodes with an interposed solid insulating barrier

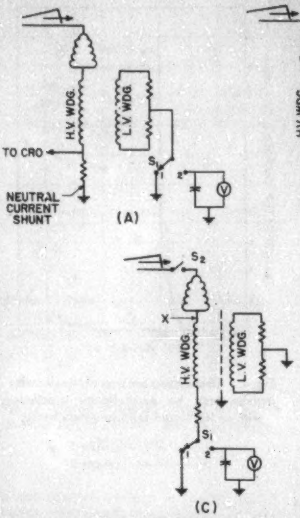


Fig. 5. Basic circuits for detecting corona in windings during impulse tests

A—No shielding between high-voltage and low-voltage windings
B—With an accessible ground shield between high-voltage and low-voltage windings
C—With an inaccessible shield between high-voltage and low-voltage windings

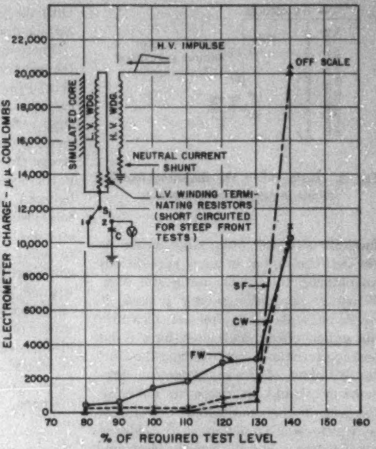


Fig. 6. Electrometer readings obtained during impulse tests on a 138-kv developmental coil stack

FW—Full-wave tests
CW—Chopped-wave tests
SF—Steep-front-wave tests

corona charge laid on the insulation for such a streamer increases with the duration of the applied voltage wave. For full-wave voltage applications the electrometer indicated corona at the same level or one level before the balanced impulse bridge detector which was also used.

For very-high-capacitance test samples the electrometer circuit of Fig. 3 may not be practical because of severe dielectric absorption and the high value of test-piece capacitance that is effectively placed across the electrometer terminals. In some such devices it may be possible to use the additional probe method of Fig. 2, and some success has been obtained in impulse tests of developmental capacitors using the capacitor case as this additional probe. More work needs to be done in this direction.

Developmental Coil Stacks Representing Transformer Windings

The detection of corona during impulse tests on a winding is particularly difficult and it was primarily for this application that the electrometer technique was developed. When the coil stack consists of two or more windings which are not electrostatically shielded from each other, the electrometer can generally be connected to the nonimpulsed winding as shown in Fig. 5(A). During application of the impulse voltage (steep front, chopped wave, or full wave) switch S_1 is in position 1, grounding the low-voltage winding through termination resistors to limit the voltage transferred to it. If desired, the voltage waves appearing on the low-voltage winding terminals may be displayed on an oscilloscope as an aid to fault detection. Immediately after application of the impulse wave, the switch is moved to position 2. The low-voltage winding then "floats" on capacitor C, and, assuming that corona is present, some of the charge induced on the low-voltage winding due to residual corona charges in the insulation structure will gradually collect on the capacitor C. This charge produces a voltage which is read with the electrometer V. It is to be noted that with this circuit the high-voltage winding is effectively grounded through both the neutral current shunt and the impulse generator discharge resistor while the electrometer is in use. This is advantageous in that charges produced by entrance bushing corona will not influence the electrometer reading.

In some transformer winding arrangements the high-voltage and low-voltage windings are effectively shielded from each other by grounded shields placed between them. For such cases no charge will be induced on the low-voltage winding by residual charges produced by high-voltage winding corona. If the ground connection to the shield is accessible, as it often is on developmental coil stacks, the circuit shown in Fig. 5(B) is applicable. With this circuit the charges induced on the ground shield are collected and read in the same manner as those induced on the low-voltage winding in the circuit described. If the ground shield is not accessible, or for other cases in which the high-voltage and low-voltage windings are shielded from each other, the circuit of Fig. 5(C) may be used. With this circuit, charge readings are made on the winding which was impulsed. Here, as before, switch S_1 is thrown to position 2 after the impulse wave has been applied, but now the impulse lead is opened at S_2 before readings are taken with the electrometer. This is necessary because in

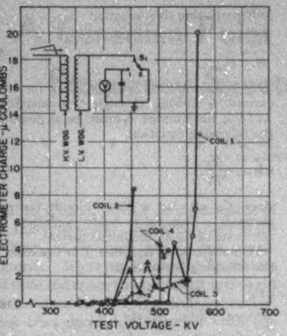


Fig. 7. Electrometer readings obtained during full-wave impulse tests on major insulation in 69-kv coil stacks

any electrometer corona detection test the portion of the winding on which induced charges are measured must "float" on top of the electrometer and its associated capacitor C (i.e., there must be no impedance path to ground except through dielectric leakage). With the circuit of Fig. 5(C) it is to be noted that the entrance bushing is also in the charge-measuring circuit. Residual charges on the bushing porcelain will, as they leak off, also contribute to the electrometer reading. Quite often the high-voltage bushing charge is small compared to that produced by winding corona and it may be neglected. For tests on developmental coil stacks made in open test tanks the switch S_2 may be placed between the bushing and winding; at position X on Fig. 5(C). It may then be arranged to ground the bushing center conductor and isolate the line end of the winding when switched to position 2, thus eliminating all interference from bushing or other circuit corona.

Fig. 6 shows the results obtained during impulse tests on a developmental 138-kv coil stack. This coil stack consisted of high-voltage and low-voltage windings placed concentrically on an electrostatically simulated core. The high-voltage winding was impulsed, and the low-voltage winding and simulated core were grounded during the tests as shown. The tests on this sample consisted of two steep-front waves, two chopped waves, and one full wave applied per level, starting at about 80% of the required test level and increasing in 10% steps to breakdown. Electrometer readings were taken after the application of each type of wave at each level. Plotted are the maximum electrometer charges versus per cent of required test level for each type of wave applied. Only gradual increases in electrometer readings were noted up through the 130% test level, and then a sharp increase occurred on application of the 140% waves. Failure of the sample over the major insulation occurred on application of the first 150% steep-front wave. The sharp increase in corona intensity just prior to the breakdown level is characteristic of this type of test sample and from it the breakdown level can often be predicted. This may prove very valuable in the future in that it may facilitate the nondestructive testing of such samples. Examination of the neutral current oscillograms which were taken for fault-detection purposes did not reveal any indication of corona.

A somewhat different application of this detection scheme is shown in Fig. 7. This test was designed to investigate the corona characteristics of the major insulation in the gap between the high- and low-voltage windings of 69-kv class transformers. The samples consisted of coil stacks made up of high- and low-voltage windings as shown with the major insulation being different for each of the four coil stacks. The impulsed high-voltage winding was short-circuited in order that all of the major insulation would be stressed equally. The low-voltage winding was solidly grounded during the impulse application, and through the electrometer after the impulse application for corona detection and evaluation. It should be noted here that an attempt was made to detect corona by the balanced impulse bridge method but it proved to be unsuccessful because of interference from high-frequency oscillations in the short-circuited loops of the high-voltage winding. The tests on each coil were discontinued before breakdown occurred in order that information correlating corona damage with corona intensity be obtained. The plotted data for coil stacks 1 and 2 again show a sharp increase in corona intensity starting at a particular test voltage, indicating imminent failure if much higher voltages were applied. The tests on coil stacks 3 and 4 were discontinued at somewhat lower corona levels to determine if damage would result from such low-intensity corona. Upon tear-down and inspection of these coil stacks it was found that only one, sample no. 1, had experienced permanent physical damage resulting from the corona discharge. The results of this set of tests gives rather convincing evidence of the potentialities of the electrometer method of corona detection as applied to the nondestructive testing of expensive high-voltage test samples.

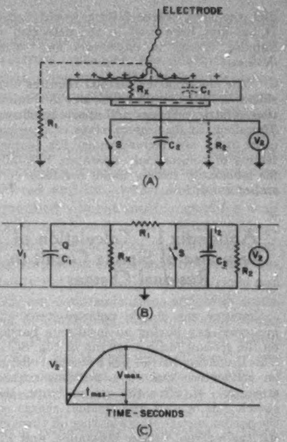


Fig. 8. A—Schematic diagram of electrometer corona detection circuit. B—Equivalent circuit of (A). C—Response of circuit (B)

R_1 = equivalent external leakage resistance to ground
R_x = internal leakage resistance of barrier
R_2 = measuring circuit leakage resistance
C_1 = equivalent capacitance of charged area
C_2 = shunt capacitor across electrometer
V_1 = potential across barrier due to corona charge Q
V_2 = voltage across electrometer
S = switch which is opened after test voltage is removed

The applications which have been presented and discussed are by no means all-inclusive and are only intended to illustrate the principles involved. The authors and their associates have used the electrometer technique on many related test samples and it is expected that there must be many other applications with which they are not familiar. It is hoped that this paper may aid in bringing such applications to light.

Conclusions

1. The electrometer method is a useful technique for indicating the occurrence and relative magnitudes of corona discharges in many electric devices. For impulse tests, it is equally effective for steep front, chopped waves, and full waves, and is not restricted by voltage magnitudes.

2. This method has been used with significant success in the laboratory testing of high-voltage coil stacks. It has been found possible to predict permanent damage before breakdown, and in many cases to pre-

dict the level where failure would occur. It has also been used in the detection of impulse corona in capacitors and other devices.

3. The electrometer method requires considerable skill in application, and in elimination of stray charges and spurious effects. Therefore, in its present state of development it is not recommended for factory testing of commercial apparatus, but only for laboratory testing under carefully controlled conditions.

Appendix I. Calculation of Response of Simple Circuit to Residual Charges

Consider the simple point-to-plane oil-immersed gap having an insulating barrier on the ground plane as shown in Fig. 8(A). This insulating barrier has a residual charge on its surface due to a previous corona streamer. If, after the corona burst, the switch S is opened, the residual charge on the barrier surface starts leaking off through R_x, the internal barrier resistance, and R_1, the external leakage resistance to ground. A little rearranging of this circuit results in Fig. 8(B) which is easily recognized as a standard impulse generator circuit. R_x can be omitted for the moment, recognizing that its effects on the electrometer reading produced by the rest of the circuit can be considered later. In Fig. 8(B) a charged capacitor C_1 (the barrier to ground plane capacitance of the charged area) has a voltage V_1 across it due to a residual charge Q on its surface. This charge Q leaks off through the external leakage resistance R_1 to C_2, charging up C_2 to some voltage V_2 indicated by the electrometer voltmeter. After reaching a maximum value, this voltage V_2 across C_2 gradually decays to zero due to leakage R_2 across the capacitor. Thus, the voltage waveshape of V_2 is approximated by Fig. 8(C). The voltage V_2 produced by such a circuit has been derived by others[7] and is shown to be

$$V_2=\frac{V_1}{2R_1C_2B}\left[\epsilon^{-(A-B)t}-\epsilon^{-(A+B)t}\right] \qquad (1)$$

where

t = time in seconds after S_1 is closed

$$A=\frac{R_1C_1+R_2(C_1+C_2)}{2R_1R_2C_1C_2}$$

$$B=\sqrt{A^2-\frac{1}{R_1R_2C_1C_2}}$$

The time required for V_2 to reach a maximum value V_m is given by

$$t_m=\frac{1}{2B}\log\frac{A+B}{A-B} \qquad (2)$$

At $t=t_m$ the maximum voltage is then

$$V_m=\frac{V_1}{2R_1C_2B}\left[\epsilon^{-(A-B)t_m}-\epsilon^{-(A+B)t_m}\right] \qquad (3)$$

and if for V_1, one substitutes its equivalent

$$V_1=\frac{Q}{C_1}$$

the equation becomes

$$V_m=\frac{Q}{2R_1C_1C_2B}\left[\epsilon^{-(A-B)t_m}-\epsilon^{-(A+B)t_m}\right] \qquad (4)$$

which shows that the maximum voltage reached by the electrometer is directly proportional to the initial charge Q left by the corona. These relations are unwieldly and unsuitable to easy calculation. However, a drastic simplification occurs if, instead of being concerned with voltage V_2 at any time t_2, one measures the rate of change of voltage dV_2/dt.

Differentiating equation 1

$$\frac{dV_2}{dt}=\frac{V_1}{2R_1C_2B}\left[-(A-B)\epsilon^{-(A-B)t}+(A+B)\epsilon^{-(A+B)t}\right] \qquad (5)$$

and at $t=0$, the initial rate of rise dV_2/dt is

$$\left.\frac{dV_2}{dt}\right|_{t=0}=\frac{V_1}{2R_1C_2B}[-A+B+A+B]$$

or

$$\left.\frac{dV_2}{dt}\right|_{t=0}=\frac{V_1}{R_1C_2} \qquad (6)$$

or in terms of initial charge Q_0 on C_1, the relation $V=Q_0/C_1$ at $t=0$ can be substituted into equation 4 and the following obtained:

$$\left.\frac{dV_2}{dt}\right|_{t=0}\quad\frac{Q_0}{R_1C_1C_2} \qquad (7)$$

Thus, if an electrometer-type meter is used which measures the rate of change of voltage across C_2 instead of voltage alone, the charge delivered is related to the rate of voltage rise by the simple relationship of equation 7. In all previous calculations, the effect of R_x, the insulation leakage resistance, has been ignored. This can now be included in the calculation. Going back to Fig. 8(B) the voltage across C_2 at any instant is given by

$$V_2=\frac{1}{C_2}\int i_2dt$$

Differentiating this equation

$$\frac{dV_2}{dt}=\frac{i_2}{C_2} \qquad (8)$$

This relation holds at any time for any current i_2 flowing through the capacitor. At $t=0$, the current i_2 must be as given in the following equation

$$i_2\big|_{t=0}=\frac{V_1}{R_1} \qquad (9)$$

since C_2 acts as a short circuit (is uncharged). This value holds regardless of the value of R_x at the initial instant.

By substituting equation 9 in equation 8 the following is obtained:

$$\left.\frac{dV_2}{dt}\right|_{t=0}=\frac{V_1}{R_1C_2}=\frac{Q_0}{R_1C_1C_2} \qquad (10)$$

which are the same as equations 6 and 7. Thus it can be stated that if, at the instant the switch S is opened in Fig. 8(A) the rate of voltage rise across C is measured, it is related to the external leakage resistance, electrometer capacitance, barrier charge, and barrier voltages as shown in equations 6 and 7. It can also be stated that it is independent of barrier leakage resistance at the initial instant.

References

1. A Method for Detecting the Ionization Point on Electrical Apparatus, G. E. Quinn. *AIEE Transactions (Electrical Engineering)*, vol. 59, Dec. 1940, pp. 680–82.

2. Development of Corona Measurements and Their Relation to the Dielectric Strength of Capacitors, R. J. Hopkins, T. R. Walters, M. E. Scoville. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1643–51.

3. Impulse Corona—Detection, Measurement of Intensity, and Damage Produced, J. H. Hagenguth, T. W. Liao. *Ibid.*, vol. 71, pt. III, Jan. 1952, pp. 461–65.

4. Detection and Study of Ionization of Industrial Dielectrics, M. F. Viale. *Revue d'Electricite et de Mecanique*, Paris, France, vol. 96, 1954, pp. 28–40.

5. Ultrasonic Detection and Location of Electric Discharges in Insulating Structures, J. G. Anderson. *AIEE Transactions*, vol. 75, pt. III, Dec. 1956, pp. 1193–98.

6. The Propagation Mechanism of Impulse Creepage Discharges Over Oil-Immersed Surfaces, J. G. Anderson, T. W. Liao. *Ibid.*, vol. 74 pt. I, May 1955, pp. 218–26.

7. Surge Phenomena (book). *Reference S/T 35*, British Electrical and Allied Industries Research Association, London, England, 1941, pp. 153–55.

8. Physical Concepts of Corona in Capacitors, J. R. Nye, W. R. Wilson. *AIEE Transactions*, vol. 72, pt. III, Aug. 1953, pp. 781–87.

A Canadian Point of View on Radio-Frequency Spectrum Management

C. M. BRANT
NONMEMBER AIEE

IT IS CUSTOMARY in a discussion of this type to start with the history of the subject as far back as one can obtain records. However, because of limitations of time and space and because it is thought to be more important to look forward, it has been decided to dispense with historical background except to mention that, since Marconi invented the tuned circuit, which required a particular frequency in cycles per second of a radio wave to be used by both transmitter and receiver, the game of the assignment of these numbers representing the frequency of the radio wave has become much more complex.

It was first suggested that this discussion be on radio licensing and later that it be concerned with radio regulation but since radio regulation is concerned with the end result of the proper assignment of numbers representing frequencies to be used by radio stations and since licensing is merely the outward documentary manifestation of this regulation, the discussion almost inevitably becomes one of Canadian spectrum management.

The radio frequency spectrum is, of course, a natural resource but it is a peculiar natural resource inasmuch as it is inconsumable; it can always be rearranged. It has, however, a top limit with regard to the number of numbers or number of frequencies which may be assigned for a given development of the art in the use of radio frequencies. Fortunately, the art is developing rapidly and the top limit is being forever pushed further up. There is another top limit as to the order of frequency or the maximum number which may be assigned. In this respect also the frontiers are continually being moved. A few years ago, the highest frequency which could be used was of the order of 100 million cycles per second; now, assignment of frequencies of the order of 10 billion cycles per second, while not yet commonplace, is being done to some extent.

To give some idea of the magnitude of the problem it should be mentioned that there are about 48,000 radio stations in service in Canada excluding broadcasting stations and mobile stations of the military services. This is five times as many as were in existence in 1948. Canada is the third or fourth (it is difficult to say which) greatest user of radio among the countries of the world.

The great problem is: 1. to arrange the orderly use of frequencies so as to provide the correct order to services able to make the most use of them, at the same time ensuring interference-free operation, and 2. to endeavor to obtain high utilization of all the frequency channels.

Two of the basic philosophies of spectrum management are:

1. To restrict the type of service for which the use of radio is permitted and so restrict the demand to fit the spectrum.
2. To permit the use of radio for any communication requirement which cannot reasonably be economically satisfied by other means.

It is the latter philosophy which the Department of Transport endeavors to follow, but examination of this statement will show that it applies no top limit to the number of frequencies which one may be called upon to assign. Obviously, since the spectrum is limited, some way must be found to accommodate the increasing number of radio stations. In Canada it is believed that the solution lies in the advances being made in techniques and that, as spectrum loading increases, advantage must be taken of newer and newer techniques which provide for greater utilization of the spectrum. These newer techniques are frequently and initially more costly to employ and the desire for their employment by the spectrum managers may appear to conflict with the manufacturing industries' desire to reduce costs in a competitive world and thereby to increase their sales of equipment. Obviously, sales can continue to increase only if spectrum space can be made available to accommodate the increased communication services represented by those sales. Therefore in this last respect (the long-term viewpoint) the interests of the manufacturing industry and those of the spectrum manager are identical. But what of the user? He requires the cheapest communications he can obtain commensurate with the speed and efficiency necessary for his purpose. A badly managed spectrum, one in which the techniques involved have not been considered and into which the services are crammed without much regard for their effect upon each other, would perhaps provide the user with the cheapest possible radio communications but would certainly not provide him with the rapid and efficient communications which he needs. As in everything else, he would obtain the poorest end result for the lowest price. If the user were allowed to employ equipment which is wasteful of spectrum space will inevitably result in a restriction of the number of users. The total requirement of the user is therefore also identical with the long-term desire of the manufacturing industry and with the desire of the spectrum manager.

The International Aspect

Before proceeding with further details of spectrum management the international aspect of radio communications should probably be discussed. Since radio radiation does not respect national boundaries, international agreements on the use of frequencies is mandatory if chaos is to be avoided. This co-ordination is arranged by the International Telecommunications Union, an organization of the United Nations and one of the oldest international organizations in existence. Conferences of member states of the International Telecommunications Union develops, manages, and modifies from time to time what is known as the International Table of Frequency Allocations and also develops international rules for the game of successfully assigning frequencies. Here a differentiation will be noted between allocation and assignment. As every professional frequency man knows, allocation is the docketing of certain frequency orders or bands for particular types of service such as aeronautical mobile, fixed services, maritime mobile and, say, broadcasting; while assignment is the selection of particular discrete frequencies or numbers for use by specified radio stations.

The last international frequency table was prepared at the Atlantic City conference of the International Telecommunications Union in 1947, and was to have superseded the Table of Cairo, 1938. However, the difficulties of implementa-

---

Paper 57-647, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 21, 1957; made available for printing April 17, 1957.

C. M. Brant is with the Radio Regulations Division, Department of Transport, Ottawa, Ont., Canada.

tion were so great that the question of the effective date of the Atlantic City Table and the date by which the new table was to become legal was left in abeyance. These difficulties were many, and not the least of them was the heavy demand by individual services such as those mentioned for bands of frequencies to meet their immediate and foreseeable future requirements. In the writer's opinion, however, the greatest single difficulty and basic problem was the lack of maneuvering space. For example, the Atlantic City conference was the first which recognized the need for exclusive bands of frequencies to be set aside for aircraft operations which had hitherto been scattered throughout the spectrum wherever frequencies could be found. Bands of frequencies were selected and the aeronautical world was to plan its own utilization within these bands. This they did very successfully, but how were they to get their bands clear for use so that they could give up frequencies which could be occupied by people endeavoring to move out of their way?

This situation is now known among frequency men as a "frequency chain reaction." It was almost impossible to work out all these chain reactions on paper, let alone to get the reaction started in fact. Implementation of the Atlantic City Table was considered during 1951 in Geneva, Switzerland, at a conference of the International Telecommunications Union known as the Extraordinary Administrative Radio Conference. It was decided to proceed on a voluntary and evolutionary basis, everyone doing his best toward implementation according to a set of rules of the game which were developed by the Conference. The Cairo Table was to continue to be the legal table for the time being, and when sufficient progress had been made in the implementation of the Atlantic City Table a short compulsory phase would be started and the Atlantic City Table would then become legal. The final phase is now at hand. Substantial progress has been made, Canada has completed about 90% of the adjustments, and an International Telecommunications Union Conference in 1959 will again review the table of frequency allocations so that this round of keeping up with developments can be started all over again.

Statutes Under Which the Administration Operates

The Administration in respect to the licensing of radio stations operates mainly under the Radio Act. This is essentially a technical act. It provides in permissive fashion for the licensing of all classes of radio stations, except those aboard ships, by the Minister of Transport. It also prescribes the qualifications in respect to the citizenship of persons who may operate radio stations in Canada and in this respect it provides for privileges for non-Canadians who are citizens of countries offering similar privileges to Canadians.

The Radio Act provides for two sets of regulations:

1. Those authorized by the Governor in Council in respect to

(*a*) Tariff of fees for licenses and examinations for certificates of proficiency in radio.

(*b*) Censorship and control of radio signals and messages in case of actual or apprehended war, rebellion, riot, or other emergency.

(*c*) Acceding to any international convention in connection with telecommunications, and issuance of any regulations necessary to carry out and make effective the terms of such convention.

2. Those authorized by the Minister of Transport in respect to

(*a*) Form and manner in which applications for licenses are made.

(*b*) Classification of stations and the technical requirements of the radio equipment to be installed, the frequencies to be used, and the nature of service to be rendered by the several classes of stations.

(*c*) Definition of different kinds of licenses and the several periods for which they may be continued in force.

(*d*) Conditions and restrictions to which the several licenses shall respectively be subject.

(*e*) Different classes of certificates of proficiency of radio operators and the class of certificate, if any, necessary to qualify a person as an operator for a coast, land, or mobile station.

(*f*) Examination of persons desiring to obtain certificates of proficiency as radio operators; qualifications in respect thereto; term of service, skill, character, etc., to be required by the applicants for such certificates.

(*g*) Secrecy of messages

(*h*) Watches, if any, to be kept by operators and number of operators, if any, to be maintained at coast, land, and mobile stations.

(*i*) Inspection of radio stations.

(*j*) Compulsion of all radio stations to receive, accept, exchange, and transmit signals and messages with such other stations and in such manner as the Minister may prescribe.

Radio stations aboard ships are licensed by the Minister of Transport under the authority of the Canada Shipping Act. This Act also prescribes the complement and class of radio stations to be carried by different categories of ships.

Broadcasting falls under the general provisions of the Radio Act and its regulations, in so far as the licensing and technical requirements of such stations are concerned. Other aspects of the control of broadcasting fall under the Broadcasting Act.

Technical Standards

The International Telecommunications Union sets minimum technical standards governing those characteristics of radio transmitting stations which affect frequency spectrum utilization. In the main, these standards are designed to take care of the effect of radio emissions of one state upon another. However, most countries set for themselves standards higher than those of the International Telecommunications Union, which is designed for spectrum economy not only with regard to the international picture but also with regard to special domestic problems. In Canada this is done by the development of technical specifications governing the suitability of radio equipment for licensing for various services. These specifications are developed by consultation between officers of the Department and the Canadian Radio Technical Planning Board, an organization representing the radio manufacturing industry and users. The specifications are intended to represent a technical level of excellence which is reasonable in the present state of the art. They are of four general types:

1. Specifications designed for standardization of techniques or methods where compatibility of radio stations in a like service is important. These apply particularly to, say, broadcasting (including television) stations and, for example, ensure that VHF (very high frequency) television receivers may be designed and built to receive any VHF television transmitting station within range, or allow a television receiver to be moved from city to city within the area using that standard and to be capable of use on the local VHF television station. Further, such standards make possible the development of large integrated mobile systems such as aeronautical mobile, maritime mobile or, say, the land mobile systems of railway companies and bus companies whereby a mobile unit may obtain service from any one of many base stations.

2. Specifications governing technical factors which have an effect upon spectrum management or frequency economy.

3. Specifications which in addition to those of item 2 also govern the reliability of radio equipment to be used for safety services such as aeronautical mobile and maritime mobile systems. In these specifications the Department is concerned with environmental effects upon reliability such as those of temperature, pressure, vibration, and humidity.

4. Fact-finding specifications which cover

mainly services or applications not already covered by detailed specifications of the type just mentioned. These are designed to put before departmental technical officers all relevant detials of the equipment or communications system proposed so that they may exercise their judgment in accordance with sound engineering practice to ensure good frequency spectrum management.

With regard to these fact-finding specifications, it is difficult for the Administration to ensure that the same kind of judgment is exercised in every case, and though it may sound alarming, this type of specification serves only as an interim measure pending the development of a standard specification which lays down clearly all the factors.

The Department recognizes that as far as possible there must eventually be a yardstick with which to measure the suitability of all radio equipment for licensing. These yardsticks can be developed only by consultation between representatives of the manufacturing industry, user organizations, and departmental officers, so as to achieve results commensurate with their common interests.

The art is developing rapidly and new techniques for the use of radio are constantly being introduced. Care must be taken not to rush into the development of standards based on insufficient or erroneous information, but at the same time, a vacuum must be avoided, so that the first applications for the use of the new techniques are not held up indefinitely. In such cases departmental officers arrange to study and discuss in some detail the objectives and aims of the sponsor of the new technique and endeavor to obtain the views of those desiring to employ it.

AMORTIZATION

Specifications of the second and third types mentioned are usually issued well in advance with a date on which they become effective for new licenses. They usually also contain a date by which existing equipment covered by existing licenses must be replaced with equipment meeting this specification. For example, certain specifications which became effective on July 1, 1956, for new licenses issued after that date require that equipment not meeting the new standard and covered by licenses issued before that date must be replaced by July 1, 1961. This is not to say that the Department considers 5 years as the amortization life of this equipment. It is merely that the specifications issued in the last 2 or 3 years by the Department are the first of this type and most of the equipment in use was several years old before that specification was issued. It is expected that future specifications superseding those now in effect will carry a longer amortization period. The period of amortization is also determined by consultation between the Department and the Canadian Radio Technical Planning Board.

LONG-TERM PLANNING FOR TECHNICAL SPECIFICATIONS

The Department is continually looking ahead with a view to ways and means to make it possible for more and more radio communication services to be accommodated in the spectrum and, obviously, the technical standards governing the suitability of equipment for licensing must be modified as the art progresses. A very good example of this type of planning is the present thinking with regard to land mobile services. Assignments of frequencies in this service are at present placed 60 kc apart. The repetition rate of frequencies of the order of, say, 150 to 170 mc (megacycles), is 54 times across the country, i.e., each frequency is assigned to approximately 54 areas in Canada so spaced that they do not interfere with each other. It will be recognized that those services close to the border of the United States must be co-ordinated with the appropriate administration of the United States. Very close co-operation exists in this field and all assignments of this order of frequency within 75 miles of the border are co-ordinated between the two countries. The art in respect of this order of frequency and in respect of land mobile equipment has now developed to the point where services could be spaced 30 kc apart, thus doubling the spectrum space available and the number of services which can be accommodated. The Department has at present under consideration the possibility of taking advantage of this development. The rechanneling of the spectrum would have to take into account existing services using equipment capable of not less than 60-kc spacing. Much of this equipment is new and meets the requirements of the last specification which became effective on July 1, 1956. It must be allowed a reasonable life before one can expect it to be discarded or modified.

It would seem that at some future date a new specification calling for 30-kc spacing should be introduced and that the services utilizing this equipment will have to be assigned frequencies with 60-kc spacing until the equipment at present in use is retired, at which time the spectrum can be rechanneled at 30 kc. This will, of course, take some years, perhaps 10 years, before the final objective can be achieved. Long before that time, 15-kc spacing will be practical. In fact, it is practical now with equipment more expensive than that presently being licensed, so that at some intermediate date within the 10 years mentioned, a specification might have to be developed calling for equipment capable of 15-kc spacing to be used with 60- and 30-kc assignments until the 30- and 60-kc equipment has been retired. By that time the cost of 15-kc equipment will have come down to a price comparable with that of present equipment. This would take probably another 5 to 6 years. By this process, the present frequency spectrum space could be made to accommodate four times the present number of radio services in about 15 to 16 years. The present rate of increase of demand for radio services for land mobile use seems to indicate that the number of services to be accommodated will be quadrupled in about this same period and could quite well increase to this number in a considerably shorter space of time.

Groups of users could perhaps work out among themselves earlier use of channel splitting. The Department would be please to receive any such co-ordinated proposals.

The Aeronautical Mobile Service

The Extraordinary Administrative Radio Conference mentioned earlier provided bands of frequencies to be used by the aeronautical mobile service exclusively for communications with aircraft. This aeronautical band space was acquired largely at the expense of fixed or point-to-point services which are now very seriously cramped for frequency space. It is unlikely that additional spectrum space can be obtained for aeronautical use by agreement in any future international conference. In the meantime, aeronautical services have increased to the point where there are not enough frequency channels to provide adequate and reliable communications for the services. It is felt that the solution lies in the development of equipment capable of providing the same communication service as at present but with less bandwidth or frequency spectrum occupancy than that utilized by the present equipment. It is believed that the aeronautical world will have to find its own solution by the use of suppressed carrier or, say, single sideband radiotelephone techniques and the rechanneling of their bands to provide for more frequency channels.

Fixed Services

The great demand for radio frequencies for fixed services is in the high-frequency portion of the spectrum between 3 and 27.5 mc to provide communications over long distances over which land-line or cable communications are not available or cannot be provided economically within reason. Each long-distance circuit requires a family of frequencies because the order of frequency varies from daytime to nighttime, from summer to winter, and from minimum to maximum solar activity. For example, a North Atlantic circuit of high frequency requires at least three frequencies for any given set of conditions, i.e., a day frequency, a night frequency, and a transition frequency to be used during intermediate periods. Over the whole year and for the complete sun-spot cycle, five or six frequencies may be required for each circuit.

The solutions for the problems of fixed services are many and all need to be used to alleviate the present congestion. A few only will be mentioned. New long underwater coaxial cables have developed which, though expensive, give very reliable service, much more reliable than that provided by radio circuits affected by ionospheric storms and aurora conditions. However, these cables have not as yet produced any alleviation of the demands for high frequencies because they are still required for stand-by and because the desire for increase communication services is outracing the laying of such cables.

Side-band suppression techniques are extensively used in high-grade, high-frequency radiotelephone services. There are, however, still too many low-grade services operating on the high-frequency spectrum, i.e., between 3 and 27.5 mc. This is obviously because the high-frequency band offers the cheapest solution for low-capacity communications circuits. If frequencies for important long-distance circuits are to be continued, more attention will have to be paid to the transferring of these low-grade circuits to VHF, UHF (ultrahigh frequency), and microwave relay systems.

Because of the congestion on frequencies of this order it is quite possible that at the next conference of the International Telecommunications Union there will be a drive to allot priorities for frequencies of this order to long-distance circuits and especially to overseas circuits. The Department, of course, believes that similar priority should be assigned to many of the circuits within Canada. At the same time, it will be extremely difficult to justify the use of high frequencies for many of the circuits still in operation in Canada. This does not, of course, mean that the Department would subscribe to a complete withdrawal of high frequencies for comparatively short-distance circuits of a few hundred miles but it is felt that many such circuits will have to take advantage of side-band suppression techniques, especially those requiring transmitter powers of 250 watts or less; and many of the shorter circuits will, as has been stated, have to be transferred to much higher frequencies.

IONOSPHERIC SCATTER

Much has been said and much has been hoped for by the use of ionospheric and tropospheric scatter techniques. A great deal of commendable research has been undertaken. However, there are serious problems which must be taken into account. Within Canada in the 30- to 50-mc band there are 14 assignments per assigned frequency. The use of ionospheric scatter is likely to cause as much loss in service by the reduction of repetition rate as it is likely to produce gain for intermediate distance circuits, say, of 600 to 1,200 miles. High transmitting powers and high antenna gains are involved in the ionospheric scatter techniques. Its proponents are always able to tell about antenna angles to half-power points but, with the powers involved, what happens beyond these half power points is of more interest. At present, there is insufficient information on this subject. Further, as is well known, the directivity is not maintained after the point of reflection. On the other hand, the technique relies on reception of very low field strengths and it is often argued that, because of this, interference to other services is not likely to be serious. However, experience has shown that, in periods of high sun-spot activity, interference to ionospheric scatter systems is quite serious. If this technique is to be used, extensive portions of the 30- to 50-mc band will have to be cleared and it is of some concern to the Department that the displacement problems, i.e., the large number of services which will have to be moved in this clearing process, does not seem to be fully understood by those who are supporting extensive use of ionospheric scatter. There is much to learn in the spectrum management aspect of the use of this technique. The use of frequencies of over 50 mc poses even more serious problems because of the possibility of interference to VHF television. As is well known, the selectivity of the average television receiver leaves much to be desired and it has already been found necessary to do two things in this respect:

1. To plan for a guard band of 2 or 3 mc in width at band edges for any assignments which are made across the board, i.e., for use anywhere in Canada.
2. To plan similar separation for local communication radio stations in the immediate vicinity of television stations using channels adjacent to the band edge.

TROPOSPHERIC SCATTER

The early remarks about the apparent bright future of ionospheric scatter should also be applied to tropospheric scatter. The distances involved are shorter but the potential capacity offered by this technique is much greater. It therefore becomes very attractive as a means of eliminating repeater stations in a conventional microwave technique, but here again serious problems arise. Of the band of 152 to 174 mc, 18 mc are assigned with a repetition rate of 54 assignments per assigned frequency. The band 470 to 890 is allocated to UHF television. Although as yet not much use has been made of this band for television in Canada, it is extensively used in the United States. It is expected that it will come into greater use in Canada as television coverage expands. All present and future UHF television assignments of the United States as well as those of Canada must be protected. It is expected that there will always be large areas of Canada probably thinly populated in which viewers will rely upon fringe area reception. This will make the protection problem even more difficult.

At one time the Department felt that since the top half of the UHF television band would probably not be required for use in Canada within the foreseeable future, because the bottom half of the band provides many more channels than are available in the VHF band, it might be possible to allocate the top half of the band to tropospheric scatter services but when one looks at the high capacity of circuits which have been proposed and the spacing of channels which would be required because of receiver limitations, among other things, it became obvious that it was impractical to proceed further with this development. Even if this were done, present and future television stations in the United States in the top half of the band would have to be protected.

There are no frequency bands allocated for this technique in the Atlantic City Table. There is an ISM (industrial, scientific, and medical) band between 890 and 940 mc and a fixed service band between 940 and 960 mc. The 940- to 960-mc band is very narrow for proper development of high-capacity systems of conventional UHF relay. It is not sufficiently wide for extensive use of tropospheric scatter. If the 890- to 940-ISM band were lumped in with the fixed service band, 940 to 960, a useful band for conventional microwave relay and tropospheric scatter would probably result. This would, of course, be resisted by those who wished to use the ISM band for its original purpose and it is not yet known if anything can be done in this direction. Neither is it known if it is practical for conventional microwave relay and tropospheric scatter to share the same bands. Splitting or compartmentalizing of the conventional microwave relay bands at 2,000, 4,000, and 6,000 mc into portions for tropospheric scatter technique and portions for conventional microwave has also been mentioned. However, many major routes in Canada using microwave relay have used very nearly the total capacity of the 4,000-mc band. The same thing may happen with the 6,000-mc band as communications in the country develop. It is probable therefore that this compartmentalization of the microwave bands will not be possible. It is said that frequencies of the order of 300 mc are the most economical for the use of tropospheric scatter. This complicates the issues still further. On the other hand, there is a band between 4,200 and 5,000 mc which is used hardly at all in Canada and this would seem appropriate for the use of tropospheric scatter. Suitable and very promising equipment has recently been developed. It should be mentioned that the problems of both tropospheric scatter and ionospheric scatter will be discussed very seriously at the 1959 conference of the International Telecommunications Union.

These comments on ionospheric scatter and tropospheric scatter are not intended to discourage the use of the technique but merely to emphasize that the problems mentioned must be considered. In fact, it is always necessary to consider the spectrum utilization problems when considering a new technique.

The Maritime Mobile Service

It will be remembered that the maritime mobile service, the oldest service to use radio, was using it not only when spark transmitters were employed but in some cases before discrete frequencies were assigned for communication. The maritime mobile radio system is highly developed from a communication organization viewpoint. It is primarily a safety service and, secondarily, it provides for public correspondence whereby any ship may send a message to any place in the world by contacting the nearest maritime mobile coast station. In some respects, however, this service, because of its traditions and because of its prior right to certain orders of frequency, has become a little outmoded.

There is, as yet, no international agreement on the use of their VHF channels, although a group of countries in Europe are advancing the idea of an international maritime mobile VHF service. At a conference in The Hague early in 1957, at which delegates of Canada and the United States were present and gave to the member countries of this conference the western-hemisphere point of view, real progress towards this end has been made. The Consultative Committee on International Radio has examined many problems associated with the development of an international maritime mobile VHF system and has suggested, among other things, the use of frequency modulation and 50-kc channel spacing in lieu of the present 100-kc channel spacing.

In Canada it is believed that a certain minimum number of VHF channels should be laid down whereby ships of any nationality might obtain service at any port anywhere in the world for emergency and calling, for port operations, for public correspondence, and for intership working.

It may be necessary to allocate more than one frequency for port operations to take care of areas where ports are adjacent to each other and within range of each other's VHF transmitters. It would seem, then, that it should be possible to reach some international agreement for the carriage of four or five VHF channels by international ships and it is hoped that subsequent developments will be such that services will be available in all the countries of the world on these channels. There are, of course, many more than four or five VHF channels assigned to the maritime mobile service, but the use of these additional channels might vary with the domestic problems of the states concerned or within regions, and it would seem that regions and states should be free to develop their own regional and domestic plans for channel frequency assignments other than the first four or five mentioned.

Another interesting development is the St. Lawrence Seaway. In these narrow navigational waters it will probably be necessary for the movement of ships to be very closely co-ordinated. This will be somewhat different from the normal concept of maritime mobile service, in which it is not mandatory, as in the aeronautical service, for vessels to report their positions. It might become mandatory in the Seaway and in the approaches to the Great Lakes. At present, medium frequencies and high frequencies are available for the route to the Great Lakes but, generally speaking, the medium frequencies provide communication service over a much greater range, with consequently a greater interference pattern than is necessary for routes such as those from the Gulf of St. Lawrence to the Great Lakes, and the Department looks forward to the time when this can be established on VHF. What has been said before about a minimum number of international channels to be carried by all international ships would probably fit very well into some system for the Seaway. Other frequencies may be provided for other purposes but at least for four or five channels mentioned would guarantee communication on VHF with vessels of all nationalities. This process of the development of more VHF utilization should in the long run alleviate congestion on the medium and high frequencies.

Recently, the Department has reexamined the question of the maritime mobile radio service mainly with regard to the medium and high-frequency service. No attempt will be made here to give the full details of the result of this reexamination but a few highlights may be given.

In the maritime radio field it is believed that the basic principle of collective safety is all-important. From this emerge two further principles:

1. In general, all vessels fitted with radio should participate on the common maritime mobile frequencies in return for the privilege of being allowed the use of radio.
2. In a given coverage area and in a given frequency band the total service (safety plus public correspondence) should be provided by one agency.

Without such principles, situations such as the following could develop:

1. Many vessels would be equipped to communicate only with their own private shore station or telephone company station and communication between vessels would be impossible in time of emergency.
2. Shore stations would provide service on a limited number of frequencies and would not have the same opportunity of communication with vessels in an emergency. Valuable time would be lost as a result of having to work through other agencies.

Public correspondence should not be allowed to jeopardize the handling of distress, safety, or urgency traffic. The public correspondence service requires the backing of an integrated fixed communication service such as is provided by the common carrier communication agencies or by, say, the Department of Transport's system.

The Canada Shipping Act and regulations specify two categories of ships fitted with radio, to both of which the basic principles just outlined apply:

1. Compulsorily fitted ships which must be fitted with radio equipment to a specified complement and to a specified technical standard.
2. Voluntarily equipped ships which may be fitted to a lower complement and technical standard.

The medium and high frequencies set aside for maritime mobile use and the organized system operating on these frequencies were intended for the use of large vessels plying recognized shipping lanes. However, these same channels have come into use by large numbers of small vessels such as fishing vessels, and are used to a great extent for communications not of a safety category and in many cases for communication with private shore stations at points not served by the public correspondence service. The volume of this kind of traffic has become so large as to deny the use of the channels for the purpose for which they were originally intended. It must be noted that operating discipline within this small vessel service is very poor. Many reports are received of profane language, failure to identify the vessels, long gossipy conversations, and the like. The larger vessels complain very seriously of their inability to establish communications because of this and because of the long calls made by small vessels in an endeavor to communicate with inadequate equipment. Of course, the small vessels also complain of interference from the higher power of larger vessels. The point must be made that poor circuit discipline contributes as much to spectrum wastage as does poor equipment. The Department has been asked to reduce the authorized power of the larger vessels.

Those concerned with the technical aspect know that in the final analysis it is the ratio of wanted signal to unwanted signal which determines effective communications and while larger vessels, usually communicating over longer distances, should not use more power than is necessary, the smaller vessels should use adequate power and equipment for their requirement especially in view of the severe antenna limitations of small vessels.

It must be recognized that the use of radio by small vessels is quite essential for their safety and for their operations but the Department believes that in general it is best carried out through the medium of bona fide general-user coast stations operated by common-carrier communication companies or by the Department of Transport. Circuit discipline is much more likely to be controlled in this way. Nevertheless, where coast station facilities do not exist or where a special requirement can be demonstrated for communications not falling within the safety category or which cannot be provided on public correspondence channels, the Department will continue to license private land stations and to license vessels to communicate on frequencies other than those provided for the proper maritime mobile service. Communications of this type will be licensed as if they were land mobile services. Nevertheless, with few exceptions, notably the case of water taxis, vessels will first be required to fit for the maritime mobile services, in the interests of collective safety, before being permitted to equip for the land mobile service.

Broadcasting

In a discussion of the special field of spectrum management there is little one can say or needs to say, except perhaps to use it to illustrate a general point. In the field of braodcasting it is necessary for the applicant's technical consultant to demonstrate to the Department, in accordance with a set of engineering rules, that any new broadcasting station, including the television station, proposed for use, can be established and operated without detriment to existing services. It is necessary for the technical consultant to select a frequency and prove to the Department that the use of this frequency is feasible. This process makes it necessary for the applicant, together with his technical consultant, to consider very carefully the position regarding existing stations before submitting a new proposal. It works well. The applicants and their technical consultants regard the technical rules set up for this purpose very seriously and one has yet to see a poor submission in this field. Even on the few occasions that such technical applications have been found unacceptable to technical officers of the Department, there has usually been a good presentation of real or alleged special circumstances surrounding the case.

The important point is that other aspects of radio communication are now becoming so complex that it may be necessary in the not distant future for applicants, particularly those proposing complex systems, to avail themselves of the services of technical consultants. In general, before this can be done, it is necessary to make available to the technical consultant comprehensive information concerning services already accommodated in the spectrum. This is now possible by virtue of an automatic punch card system which has been set up in Ottawa. It has always been available in respect to broadcasting stations which, of course, are small in number as compared with communication stations. In regard to communication services the situation is frequently difficult. Applications are received which are not considered well founded, but which should not be turned down out of hand, and in which it is felt that some endeavor should be made to assist the applicant to develop his proposal into a more feasible one. The Department often finds itself in a position of endeavoring to assist the applicant against his will because to him it is necessary only that it approve his original application. The efforts of the Department are therefore often mistaken for "red tape" and as a roundabout method of turning down the proposal.

Conclusions

The term "spectrum or frequency manager" has been used as if this so-called frequency manager were an individual. This is, of course, not the case. In using this terminology one refers to a fairly complex organization consisting of the Radio Regulations Division of the Telecommunications Branch of the Department of Transport in Ottawa, 6 district offices and 17 field offices charged with the responsibility for maintaining order in the radio frequency spectrum. Nevertheless, the term "spectrum manager" will be used here.

No attempt has been made to deal exhaustively with any of the services mentioned, but each has been used to illustrate some point of spectrum management.

An attempt has been made to demonstrate that the long-term interests of the manufacturers of radio equipment, of the users of radio equipment, and of the spectrum manager, are one and the same thing. However, a good deal of co-ordination is necessary. This co-ordination need not be lengthy, formal, and time-consuming. A system of co-ordination should not be developed whereby each cross section of the manufacturing industry or each cross section of the users feels that it is called upon to press very hard for the satisfaction of its own individual demands for spectrum space at the expense of other groups. Furthermore, the spectrum managers should not be allowed to develop "ivory-tower" spectrum management ideology without regard to its economic effect upon the manufacturing industry or the users. By a free interchange of opinions, aims, and objectives on an informal basis among manufacturers, users, and spectrum managers, the desired result should be achievable. For this reason, industry representatives are encouraged thoughts to drop in on representatives of the Department from time to time to explain developments and their own thinking, and to obtain from the spectrum managers their thoughts and plans for the future. It is hoped that all users will take an interest in their medium, and that many of the larger users will consider it sufficiently serious to engage on their staffs professional radio men, not only to help them to produce the best economical plan consistent with good frequency utilization, but also to represent their interest in the field of spectrum utilization planning. In other words, if the mutual benefits of a long-term result are recognized, each participant should be prepared, in some cases, to give ground on what might appear to be a quick expedient short-term solution; to maintain the necessary informality; to be prepared to attend all sorts of meetings, preferably in small groups; and to discuss freely the mutual problems.

The result will be a well-managed spectrum, a thriving manufacturing industry, and satisfied users.

A 14-Channel Displacement Measuring Device Utilizing Magnetic and Paper Tape Recording

G. R. PARTRIDGE
ASSOCIATE MEMBER AIEE

J. W. DUNKIN
NONMEMBER AIEE

R. L. ANDERSON
MEMBER AIEE

R.C. GELDMACHER
NONMEMBER AIEE

ONE of the most pressing problems in modern life is the construction and maintenance of an adequate system of highways. It is evident that the need is great for more new highways and, equally important, they must be constructed in the most durable manner possible.

Theoretical analyses of the response of pavement-earth systems to vehicular loads have been undertaken in an effort to determine the relationship between stresses, strains, damping, and elastic constants and the deflection of the pavement surface.[1] From these studies, a great deal has been learned regarding the comparative merits of different subgrade construction methods. To obtain experimental as well as theoretical information, a series of measurements was proposed on highways of different types. The instrument described in this paper was designed expressly for this purpose. It is not by any means restricted to this application, but much of the design philosophy was colored by the fact that the equipment would be installed in the field rather than the laboratory.

The instrument is constructed in the form of two compatible 7-channel measuring devices which can, if desired, be used as a single 14-channel unit. Recording may be by means of either magnetic tape or paper chart. The first makes possible continuous long-period recording in a form suitable for electrical data-processing; the latter, short-period, directly observable recording suitable for exploratory observations. The instrument has received extensive use in one or the other of its modes of operation and its flexibility is a great asset in the field. It has met the stringent requirements of linearity, stability, sensitivity, and dependability imposed by the application for which the instrument was intended.

The system as shown in Fig. 1 is being used to record 14 simultaneous events on magnetic tape. Fig. 2 shows one of the two identical units into which the system may be separated. Each of these units can be used to drive seven 1,500-ohm recording galvanometers.

Over-All System

The principal objectives to be met in the system design were:

1. Precision.
2. Portability.
3. Provision for 14 channels of continuous recording over periods of an hour or more.
4. Provision for electrical data processing.
5. Provision for directly observable measurements in at least two widely separated field locations.

Of the two modes of operation, that of recording on tape meets the required objectives of playback into an electric data-processing system and multichannel long-period recording. The separable 7-channel units using galvonometer recording provide two independent stations of measurement with immediately interpretable observations suitable for exploratory work.

The sensitivity of the system is prescribed on the basis of preliminary (unpublished) studies. It was found that a 10,000-pound wheel load applied to a concrete pavement will produce a maximum deflection on the order of 0.010 inch and will produce no significant deflection at distances greater than 12 feet from the point of load application. In view of these facts, the Schaevitz *040 M-L* linear variable differential transformer is used as the principal transducer, and the system is designed to match it. For further flexibility, driving voltages and attenuators are provided to match the instrument to the Schaevitz *125 S-L* differential transformer.

Paper 57-662, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted November 7, 1956; made available for printing June 3, 1957.

G. R. PARTRIDGE is with the Raytheon Manufacturing Company, Wayland, Mass., and J. W. DUNKIN, R. L. ANDERSON, and R. C. GELDMACHER are with the Division of Engineering Sciences, Purdue University, Lafayette, Ind.

The instrument described was designed and developed as one activity of a Joint Highway Research Project of the School of Civil Engineering of Purdue University, headed by K. B. Woods. The project sponsors were the Bureau of Public Roads, represented by Harold Allen and L. W. Teller, and the State of Indiana, represented by W. T. Spencer. The School of Civil Engineering of Purdue University was represented by H. L. Michael, Assistant Director of the Joint Highway Research Project. The technical activities of the project were directed by members of the staff of the Division of Engineering Sciences of Purdue University, headed by Paul F. Chenea; R. C. Geldmacher was project leader. The authors wish to express their appreciation to Carlos de Andrade, Allan Knoche, Loren Bedwell, Donald Wood, Peter Hsieh, John Brownlee, and John Keller, who, as student laboratory assistants, assembled and tested the instrument and contributed to its design and development.

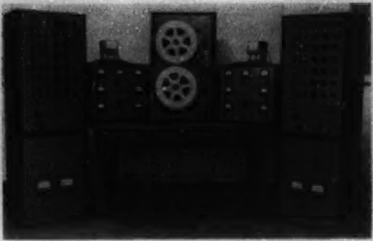


Fig. 1 (left). Fourteen-channel system recording on magnetic tape

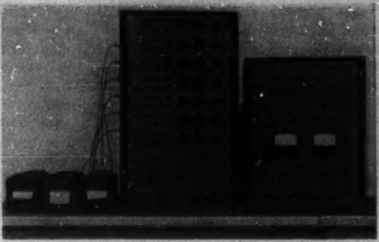


Fig. 2 (right). One of two identical 7-channel units being used to drive galvanometers for chart paper recording

The bandwidth is prescribed on the basis of the maximum expected vehicle speed and the range of influence of the loading. Assuming vehicle speeds of 70 miles per hour or less and a range of load influence of 12 feet, the shortest duration of a measured event is found to be about 1/2 second and, being a relatively smooth time function, has no significant harmonic components over 50 cps (cycles per second). Consequently the system is designed to respond to the band of frequencies which covers the range from zero to 50 cps.

In the early stages of development of the instrument, considerable effort was expended in attempting to record an AM (amplitude modulation) signal on tape. The advantages inherent in the simplicity of such an approach are quite apparent. However, observations made while introducing an amplitude-modulated signal to the tape showed that a heavy concentration of noise developed in the region under 30 cps, making amplitude modulation impractical for the desired application. At present, even though some manufacturers indicate their ability to provide tape with very low noise level, the best solution seems to be FM (frequency modulation), and the system was finally designed as an FM device, employing a deviation of ±1 kc about a midfrequency of 3.5 kc.

Fig. 3 illustrates the general arrangement. A 2-kc oscillator of the Wein bridge type and associated power amplifiers provide the signal to power the transducers. The 30-kc oscillator, also a Wein bridge, was originally used to provide recording bias for the tape when AM recording was attempted. In the FM system the output of this oscillator is rectified to give grid bias for the transducer power amplifiers. This voltage is much easier to filter than a rectified 60-cycle voltage and, as long as it was available, it was decided to take advantage of it in this way.

The outputs of the transducers are suppressed-carrier AM signals,[2] making it necessary to restore a carrier. Consequently, the input to the recording amplifier provides for adding a 2-kc reference to the transducer signal, and both the magnitude and phase of the reference are adjustable over a wide range. After the carrier is restored, the AM signal in the recording amplifier is either attenuated or amplified to a definite level for operating the recorders. If tape recording is employed, the signal is demodulated to recover the envelope, which varies the bias on a multivibrator to produce square waves of varying frequency and constant amplitude. These are recorded on the tape. When an ink record is to be made directly, the demodulated AM envelope is amplified at once and used to drive the recording galvanometers.

As a result of unsatisfactory experience with AM tape recording, the playback amplifier is designed to handle FM signals only. These are picked off by the tape recorder playback head, amplified, and clipped to square waves, which then operate a 1-shot multivibrator. The average voltage at either plate of the multivibrator is a function of the times per second that the plate is "on," and therefore a direct voltage is developed which is a linear function of the frequency of the FM signal at that moment. This direct voltage is limited to protect against possible surges, and then amplified to the level necessary to drive the galvanometers. (The limiting action obviously occurs only in case of an overload.)

The diagram of Fig. 3 does not include power supplies, which are conventional for the most part. The requirement of field use demands a gasoline-engine-driven alternator to generate 115-volt 60-cycle power for the system. If commercial a-c power is available at a particular site, this is used instead. Plate power supplies are commercially purchased, electronically regulated units. The filament supply consists of a 60-cycle induction motor driving a self-excited d-c generator of the

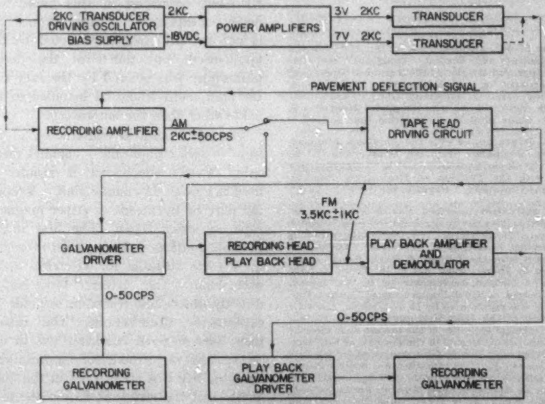


Fig. 3. Block diagram of measuring system

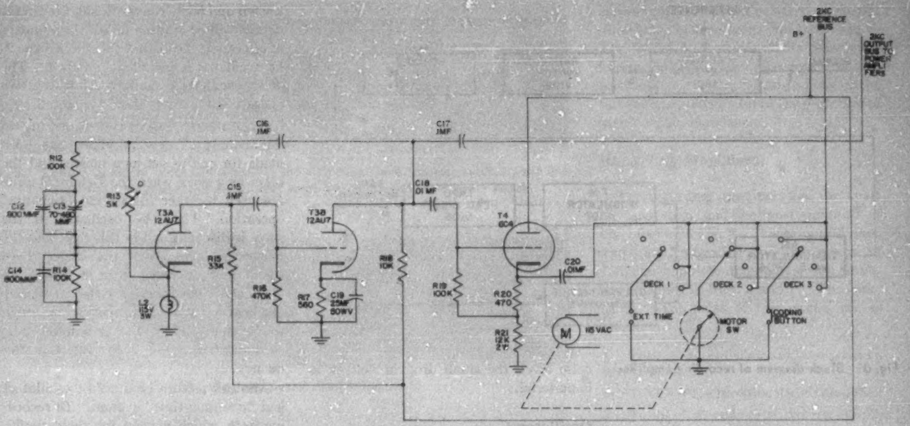

C16
.1MF
C17
.1MF
R12
100K
C12
800 MMF
C13
70-480
MMF
C14
800MMF
R14
100K
R13
5K
D
T3A
12AU7
L2
115V
3W
C15
.1MF
R15
33K
T3B
12AU7
R16
470K
R17
560
C19
25MF
50WV
C18
.01 MF
R18
10K
R19
100K
T4
6C4
R20
470
R21
12K
2W
C20
.01MF
M
115 VAC
DECK 1
DECK 2
DECK 3
EXT TIME
MOTOR
SW
CODING
BUTTON
B+
2KC
REFERENCE
BUS
2KC
OUTPUT
BUS TO
POWER
AMPLI
FIERS

Fig. 4. The 2-kc carrier frequency oscillator

type used in automobile electric systems. The weight of this assembly is approximately the same as that of a transformer that could deliver some 40 amperes of filament current, and gives the advantage of hum-free d-c heating power. This is especially important in the low-level recording amplifier stages where rectified filament power is beneficial in any event.

The circuits of the monitors for the magnetic-tape-recording level consist of cathode follower isolating stages to which are connected d-c meters.

A word now regarding the choice of frequencies. If AM recording becomes feasible, a 2-kc carrier frequency is the optimum. This is because the voltage induced in the playback head is proportional to $d\phi/dt$, where ϕ is the number of lines of flux stored on the tape. The higher the frequency, the more lines are found in a given length of tape, and the greater the voltage induced in the reproducer. On the other hand, too high a frequency results in crowding the magnetized areas together, producing a demagnetization between adjacent regions, and reducing the output.

With FM recording, a moderately large deviation is necessary in order to obtain adequate amplitude upon demodulating. A deviation of ±1 kc is difficult to obtain with a carrier frequency of 2 kc. Therefore, the center frequency is 3.5 kc for the FM system. This costs a few decibels in voltage developed at the reproducer. However, demodulation is not a function of the amplitude in an FM system. As long as the recorded signal is well over the

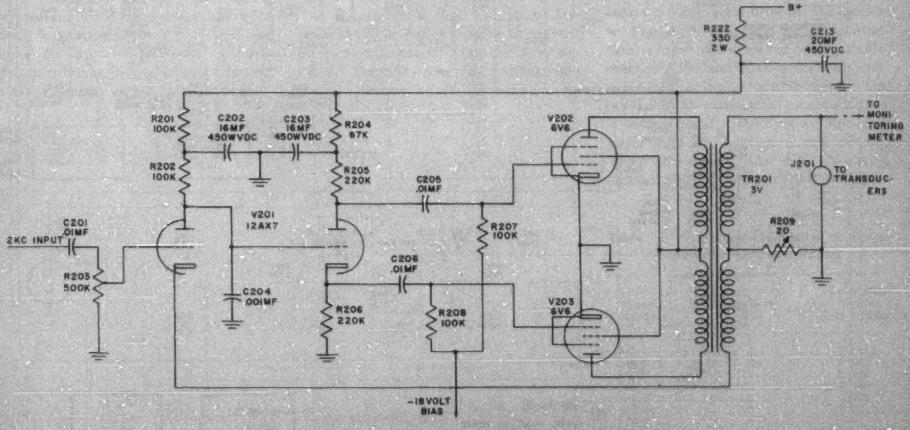

R222
330
2 W
B+
C213
20MF
450VDC
R201
100K
C202
16MF
450WVDC
C203
16MF
450WVDC
R204
67K
R202
100K
R205
220K
C205
.01MF
V202
6V6
TO
MONI-
TORING
METER
J201
TO
TRANSDUC-
ERS
TR201
3V
C201
.01MF
2KC INPUT
V201
12AX7
R207
100K
R209
20
R203
500K
C204
.001MF
C206
.01MF
R206
220K
R208
100K
V203
6V6
-18 VOLT
BIAS

Fig. 5. Power amplifier for driving the differential transformers

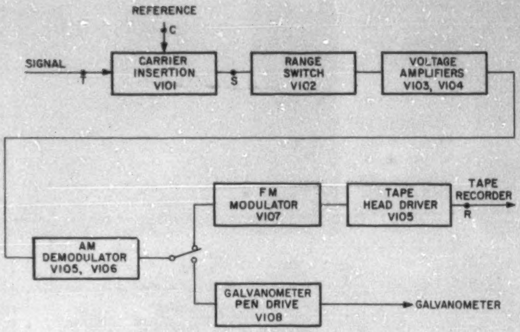

REFERENCE
C
SIGNAL
T
CARRIER INSERTION V101
S
RANGE SWITCH V102
VOLTAGE AMPLIFIERS V103, V104
FM MODULATOR V107
TAPE HEAD DRIVER V105
TAPE RECORDER
R
AM DEMODULATOR V105, V106
GALVANOMETER PEN DRIVE V108
GALVANOMETER

Fig. 6. Block diagram of recording amplifier

noise level, the small drop in voltage is immaterial.

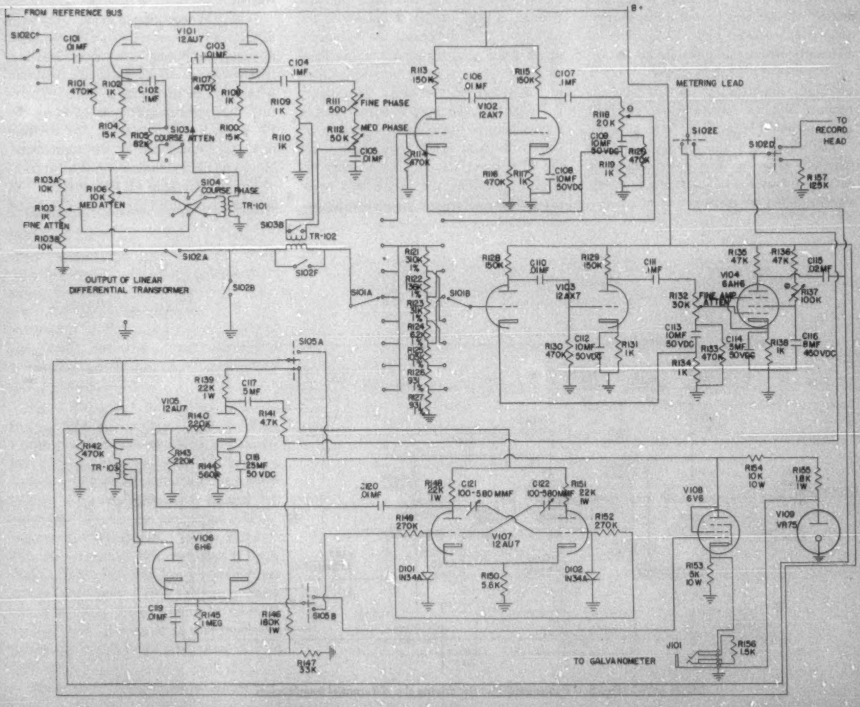
Oscillators[3]

The oscillator circuits, as noted, are of the Wein bridge type. Since this is well known and fairly conventional, the circuit details will be omitted here. As a matter of reference, the circuit diagram for the 2-kc oscillator is included in Fig. 4. The 30-kv oscillator is similar. However, two operational features should be noted.

Since the equipment is designed in two duplicate packages, provision has been made for one to act as a master and the other as a slave when the 7-channel units are used together to provide 14-channel operation. The 2-kc oscillator in the slave is disconnected in this case, and the master provides both transducer drive signal and reference voltage for all 14 channels. This prevents the possibility of a beat frequency appearing anywhere in the system should the individual 2-kc oscillators be a few cycles apart in frequency.

Another feature of the 2-kc oscillators is a "common-time" system. In recording data which depend for their significance on exact correlation of time among all measuring points, it is necessary to include timing marks that will appear

Fig. 7 (below). Complete recording amplifier for one channel


FROM REFERENCE BUS
B+
S102C
C101 .01MF
V101 12AU7
C103 .01MF
C104 .1MF
R101 470K
R102 1K
C102 .1MF
R107 470K
R108 1K
R104 15K
R105 82K
S103A COURSE ATTEN
R100 15K
R109 1K
R110 1K
R111 500
FINE PHASE
R112 50K
MED PHASE
C105 .01MF
R113 150K
C106 .01MF
V102 12AX7
R115 150K
C107 .1MF
R114 470K
R116 470K
R117 1K
C108 10MF 50VDC
R118 20K
C109 10MF 50VDC
R120 470K
R119 1K
METERING LEAD
S102E
S102D
TO RECORD HEAD
R157 125K
R103A 10K
R106 10K MED ATTEN
S104 COURSE PHASE
TR-101
R103 1K FINE ATTEN
R103B 10K
S103B
TR-102
S102A
S102F
OUTPUT OF LINEAR DIFFERENTIAL TRANSFORMER
S102B
S101A
R121 310K 1%
R122 135K 1%
R123 31K 1%
R124 6.2K 1%
R125
R126 931 1%
R127 931 1%
S101B
R128 150K
C110 .01MF
R129 150K
C111 .1MF
V103 12AX7
R130 470K
C112 10MF 50VDC
R131 1K
R132 30K FINE AMP ATTEN
C113 10MF 50VDC
R133 470K
R134 1K
C114 5MF 50VDC
R135 47K
R136 47K
C115 .02MF
V104 6AH6
R137 100K
R138 1K
C116 8MF 450VDC
S105A
R139 22K 1W
C117 5MF
V105 12AU7
R140 220K
R141 4.7K
R142 470K
TR-103
R143 220K
R144 560K
C118 25MF 50 VDC
C120 .01MF
R148 22K 1W
C121 100-580 MMF
C122 100-580 MMF
R151 22K 1W
R149 270K
R152 270K
V107 12AU7
D101 1N34A
D102 1N34A
R150 5.6K
R154 10K 10W
R155 1.8K 1W
V108 6V6
V109 VR75
R153 5K 10W
V106 6H6
C119 .01MF
R145 1MEG
R146 180K 1W
S105B
R147 33K
J101
R156 1.5K
TO GALVANOMETER

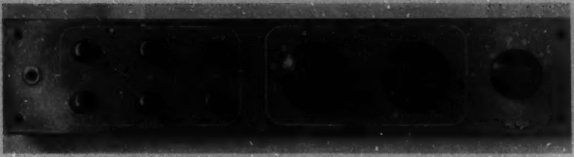


Fig. 8. Control panel of recording amplifier showing input jack, coarse, medium, and fine controls of carrier phase and amplitude, amplifier attenuators, and meter switch

Fig. 9. Metering and power distribution panel

simultaneously on the records from all channels. A small 1-rps (revolution per second) synchronous motor drives a switch with a short-circuiting contact. Short-circuiting the reference bus for about 10 milliseconds per second introduces a transient into all channels simultaneously, causing a spike to appear in the recordings. The motor-driven switch is paralleled by a pair of binding posts for connecting an external switch if a time base of other than 1 second is desired.

There is also provision for short-circuiting the reference bus with a push button. This is very useful in coding a particular run of data. For example, if test number 25 is about to be made, the button is pushed twice and after a brief pause, five times more. To prevent the timing switch from introducing pulses during the coding operation, a switch is provided giving the operator a choice of push-button coding, 1-second interval timing, external timing, or all three at once.

Power Amplifiers

As stated earlier, the transducers for which this equipment was designed are Schaevitz linear differential transformers, models *040 M-L* and *125 S-L*. The former is the "work horse" of the system, while the latter is used only where relatively large deflections, of the order of 1/10 inch, are expected. The type *040 M-L* requires relatively little input power, and one power amplifier is capable of furnishing 3 volts to 7 of these transducers in parallel. The type *125 S-L* requires 7 volts input and appreciably more power. A power amplifier will drive just one *125 S-L* transducer. (For reference purposes Fig. 5 is included, showing the circuit diagram for the power amplifier.)

It is imperative that the output impedances be as low as possible to prevent interaction from one transducer to another. For this reason, both voltage and current feedback are employed.[4] By careful adjustment of the latter, an output impedance of zero ohms can be obtained. The feedback produces all its other well-known benefits, such as freedom from distortion and a better signal-to-noise ratio, but these are incidental to the objective of zero output impedance.

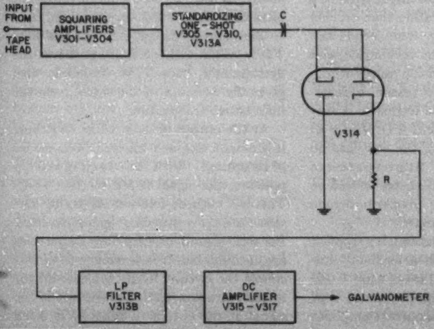


Fig. 10. Block diagram of tape playback system

Recording Amplifiers

The recording amplifiers are the heart of the system, and the most elaborate of the blocks in Fig. 3. There are 14 such amplifiers, separated into two groups of 7. The complete diagram for one amplifier is shown in Figs. 6 and 7. The functions of the parts of this circuit are as follows:

1. Tube *V101* and associated components, carrier insertion.
2. Tube *V102*, a precision decade amplifier.
3. Tubes *V103* and *V104*, voltage amplifiers.
4. Tubes *V106* and the left half of *V105*, demodulator.
5. Tube *V108*, galvanometer pen drive.
6. Tube *V107*, frequency modulator.
7. The right half of *V105*, tape-recording head driver.

Most of these functions will be examined in more detail. The carrier insertion starts with the left-hand half of tube *V101*. A cathode follower is used in this case to avoid loading the reference signal bus. There may be as many as 14 channels connected to this bus, so a high input impedance to each is very necessary. The output of this cathode follower goes first through a resistor which may be switched in or out to provide a coarse attenuation control for the carrier; see Fig. 8. Medium and vernier attenuation controls are achieved with potentiometers *R103* and *R106*, which are connected in a bridge arrangement. The output of this bridge is connected to a reversing switch, providing a 180-degree phase shift, and labelled "coarse phase." The right half of tube *V101* is also a cathode follower, used in this case to pro-

Table I. Transducer-Plus-Carrier Voltages Expected at Output of TR102

Type of Transducer	Displacement, Mils	Signal Plus Carrier, Rms Voltage
040 M-L	0.1	0.0012
040 M-L	1	0.012
040 M-L	10	0.120
040 M-L	40	0.480
125 S-L	1	0.032
125 S-L	10	0.32
135 S-L	100	3.2
125 S-L	200	6.4

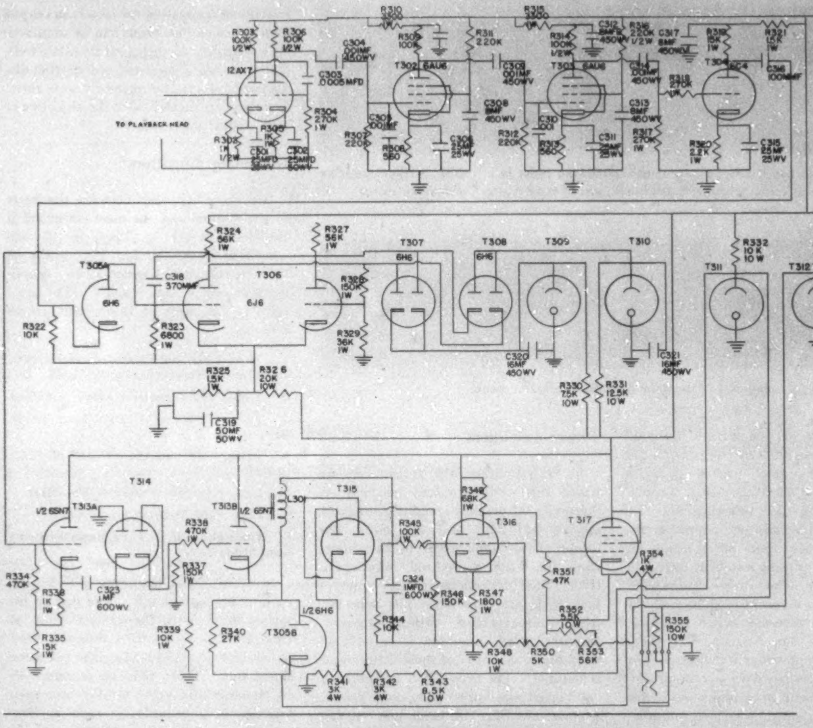

TO PLAYBACK HEAD
12AX7
R303 100K 1/2W
R306 100K 1/2W
R302 1K 1/2W
R305 1K 1W
C301 25MFD 25WV
C302 25MFD 50WV
R304 270K 1W
C303 .0005MFD
C304 .001MF 450WV
R310 3300 1W
R309 100K
T302 6AU6
C305 .001MF
R307 220K
R308 560
C306 25MF 25WV
R311 220K
C308 8MF 450WV
C309 .001MF 450WV
R312 220K
C310 .001
R313 560
C311 25MF 25WV
R315 3300 1W
R314 100K 1/2W
T303 6AU6
C312 8MF 450WV
R316 220K 1/2W
C314 .001MF 450WV
C313 8MF 450WV
R317 270K 1W
C317 8MF 450WV
R318 270K 1W
R319 15K 1W
T304 6C4
R320 2.2K 1W
C315 25MF 25WV
R321 15K 1W
C316 100MMF
R324 56K 1W
R327 56K 1W
T305A
6H6
C318 370MMF
T306
6J6
R322 10K
R323 6800 1W
R328 150K 1W
R329 36K 1W
R325 15K 1W
R326 20K 10W
C319 50MF 50WV
T307
6H6
T308
6H6
T309
T310
C320 16MF 450WV
C321 16MF 450WV
R330 7.5K 10W
R331 12.5K 10W
R332 10K 10W
T311
R333 5.5K 10W
T312
C322 16MFD 450WV
T314
1/2 6SN7
T313A
T313B
1/2 6SN7
R334 470K
R336 1K 1W
R335 15K 1W
C323 1MF 600WV
R338 470K 1W
R337 150K 1W
R339 10K 1W
L301
R340 27K
1/2 6H6
T305B
T315
R345 100K 1W
C324 1MFD 600WV
R344 10K
R346 150K
R349 68K 1W
T316
R347 1800 1W
T317
R351 47K
R354 1K 4W
R352 5.5K 10W
R355 150K 10W
R348 10K 1W
R350 5K
R353 56K
R341 3K 4W
R342 3K 4W
R343 8.5K 10W

Fig. 11. Circuit diagram of playback system

vide a low output impedance for driving the resistance-capacitance phase shift network immediately to the right of the tube. The carrier, after being properly adjusted in phase and amplitude, is added to the signal at transformer *TR102*. The sum is then either attenuated or amplified as required to yield the proper voltage for the remainder of the recording amplifier. Table I gives the voltages expected for given displacements with the two types of transducers.

The attenuator, consisting of the precision decade amplifier and resistors *R121* through *R127*, is marked on the front panel in terms of the transducer model and range rather than the voltage to be amplified. For this reason, the various taps appear somewhat disordered as shown on the circuit diagram. The design of the attenuator is such as to reduce all voltages to a maximum expected value of 12 millivolts rms before applying them to tube *V103*. This requires the use of an extra amplifier, tube *V102*, for the most sensitive range, i.e., the *040 M-L* transducer with an 0.1-mil displacement. This seems a better plan than reducing all signals to 1.2 millivolts, since noise might then be a problem on all ranges rather than on only the most sensitive one. The decade amplifier must, of course, be very stable, as it forms a part of the precision range-switching assembly. For this reason, a generous amount of negative feedback is provided, and the gain control (*R118*) is so arranged that adjusting the gain has no effect on the loop gain. Any references to the range-switching attenuator should be taken as including the precision decade amplifier for 0.1-mil ranges.

The output of the attenuator is applied to tube *V103*, which is a feedback stabilized stage similar to that of tube *V102*.

The next stage, a high-perveance type *6AH6*, may be used as a current amplifier to drive the tape-recording heads directly if AM recording should become feasible. In the present FM system, this stage is simply another voltage amplifier. Its output is taken to the metering panel, shown in Fig. 9, where the level of the recording signal may be monitored, and to the input of the left-hand half of tube *V105*. This cathode follower drives the demodulator, tube *V106*, which in turn gives the envelope of the signal from the differential transformer.

At the output of tube *V106*, switching is provided to give a choice of two modes of recording. With the switch in the top position, the signal is fed to tube *V108*. This is a cathode follower supplying current to the recording galvanometers. Because signals of very low frequencies are encountered, it is necessary to direct-couple the circuits from the demodulator to the galvanometer. This results in the signal at the output of tube *V108* being about 75 volts above ground potential. This voltage is balanced by 75 volts from tube *V109*. Consequently, the net voltage across the galvanometer is the signal component alone.

Fig. 12. Sample record showing deflection of pavement as a function of time, with truck driven over at creep speed. Note small deflection caused by front wheels, and larger deflection as rear wheels pass by. Short vertical marks are timing pips at 1-second intervals. Maximum deflection 5.2 milli-inches

If recording on magnetic tape is desired switch *S105B* is in the lower position. This applies the output of the demodulator to the grids of the multivibrator, tube *V107*. The frequency of this multivibrator is nominally 3.5 kc, but it is varied over a range of ±1,000 cps by changing the direct voltage on the grids. A variable-frequency square wave then appears at the input to the right-hand side of tube *V105*. The series grid resistor *R140* clips the signal from the multivibrator and helps insure a good square wave output from this stage. The square wave of current in the recorder head is 12 milliamperes, peak to peak. This is sufficient to saturate the tape, and eliminates the need for a recording bias.

Playback System[5]

As pointed out in discussing the recording amplifier, the galvanometers may be driven directly without the intermediate step of recording data on tape. However, a playback system is required when data taken on tape are to be converted to a pen-and-ink chart display. This playback system circuitry is shown in Figs. 10 and 11.

It will be noted that tubes are arranged in three horizontal rows. The top row consists of amplifiers, the middle row of a 1-shot multivibrator and clamping circuits, and the bottom row of demodulating circuits that convert the output of the multivibrator into current to drive the galvanometers.

With the tape fully saturated, an output voltage of approximately 100 microvolts, peak to peak, is expected. This voltage is applied to the input of tube *V301*, and amplified by this tube, and by tubes *V302* and *V303*, to obtain a total gain of approximately 120 decibels. The output at the plate of tube *V303* is about 100 volts, peak to peak, and very nearly square. This voltage is clipped top and bottom by tube *V304*, and differentiated by *C316* and *R322*, resulting in a series of narrow pulses at the cathode of *V305A*. This diode admits only the negative pulses, which trigger the 1-shot multivibrator tube *V306*. Diodes *V307* and *V308* clamp the anodes of the multivibrator at specified levels of 75 volts and 150 volts. This is done to help guard against changes in the width of the multivibrator output pulse if operating potentials should vary. Therefore, the input to cathode follower tube *V313A* consists of square pulses of standard amplitude and width with a repetition rate equal to the frequency of the signal on the magnetic tape at that instant, but approximately 3.5 kc.

The conversion of these standardized square pulses into direct current occurs in *C323* and tube *V314*. For each pulse, *C323* will charge through the right-hand side of the tube and discharge through the left-hand side. Therefore the direct current flowing in the right-hand side of tube *V313B* is a function of the frequency on the magnetic tape. The demodulation process is complete for all practical purposes at this point. However, the pulsating direct current is filtered by *L301* and *C324* to remove the (nominal) 3.5-kc components and their harmonics. Tube *V315* acts as a limiter to prevent the direct current developed at this point from overdriving the rest of the system and damaging the galvanometers. For signals of the size ordinarily encountered, the limiter would not be brought into play at all.

After limiting, direct-coupled amplifiers *V316* and *V317* bring the signal level up to the value required to operate the galvanometers. Since direct coupling to the cathode follower *V317* will place one lead of the galvanometer appreciably higher than ground potential, the other lead must be returned to a positive voltage also; in this case, the anode of tube *V312*. Otherwise, the potential across the galvanometer would drive the pen completely off scale, and probably burn out the coil as well.

The tape drive mechanism is fitted with two sets of heads, making possible simultaneous recording and playback. As a result, channel response can be accurately and quickly checked.

Conclusions

The instrument meets the requirements necessary for the precise measurement of deflections in concrete pavements resulting from moving vehicular loads. Such deflections have maximum amplitudes in the range from 1 milli-inch to 10 milli-inches and harmonic components in the range from 0 to 50 cps. A photograph of two sample records may be seen in Fig. 12. It will be noticed that low-amplitude oscillations appear on each record near the point of maximum deflection. These oscillations occur at the dominant natural frequency of the suspension system of the loaded truck and represent the forced response of the pavement resulting from the truck passing at 2 miles per hour over a set of lead wires.

The reproducibility of field measurement was observed over a period of 3 hours and a representative variation in maximum amplitudes was found to have a distribution for which the standard deviation was 0.13 about a mean value of 6.5. This deviation represents the combined effects of all sources of variation such as human errors in calibration and data reduction, pavement drift, vehicle control, and instrument drift. Studies of sources of variances show that slight changes in lateral position of the vehicle cause deviations of the order stated. It is therefore concluded that instrument errors fall well within the 0.13 deviation.

The use of the instrument is not limited to pavement deflection measurements. It could be adapted easily to the measurement of displacements occurring in buildings, bridges, and other structures.

The alternate modes of operation permit rapid exploratory measurements or long-period recording and electrical data-processing.

References

1. SUBGRADE SUPPORT CHARACTERISTICS, R. L. Anderson, J. W. Dunkin, R. C. Geldmacher, M. E. Harr, G. R. Partridge, L. E. Wood. Joint Highway Research Project, Purdue University, Lafayette, Ind., 1957.

2. THE STRAIN GAGE PRIMER (book), C. C. Perry, H. R. Lissner. McGraw-Hill Book Company, Inc., New York, N. Y., 1955. (Discusses in particular suppressed carrier systems as found in instrumentation applications).

3. RADIO ENGINEERING (book), F. E. Terman. McGraw-Hill Book Company, Inc., third edition, 1947, pp. 436–37.

4. *Ibid.*, p. 325.

5. A DIRECT-READING FREQUENCY METER FOR THE AUDIO AND SUPERSONIC RANGES, H. J. Reich, R. L. Ungvary. *Review of Scientific Instruments*, New York, N. Y., vol. 19, Jan. 1948, p. 43.

The Elimination of Motor-Generator Sets as Transient Buffers for SAGE Computers

J. J. GANO
MEMBER AIEE

Synopsis: Motor-generator sets are frequently used as buffers to isolate electronic equipment from severe power transients which may cause misoperation. An investigation of voltage tolerances for digital computers, performance of d-c supplies, and probabilities of fault occurrences indicates that if the power is generated on the site of the electronic equipment, motor-generator sets are not necessary to meet the critical requirements of the digital computers employed in the SAGE (Semi-Automatic Ground Environment) air-defense system.

BECAUSE of the high initial cost of large electronic systems, their malfunction must be kept at a minimum. Among the many reliability aids adopted is the use of motor-generator sets in the power-conversion section to isolate the electronic equipment from severe power transients which may cause misoperation. These transients are caused by disturbances in the distribution system, either in the public utility network or within the plant. The motor-generator sets also perform other functions. For tube reliability the voltage of the generator supplying the heaters is closely regulated to ±1% and the regulating equipment is adapted to increase the heater voltage linearly from zero to full voltage in one minute and the reverse for voltage removal. The regulation and gradual application of the heater voltage may be obtained by other methods. Induction regulators, saturable reactors, constant-voltage transformers, and motor-operated adjustable transformers regulate the alternating voltage very nicely. Impedance elements, inserted in series with the heaters when the circuit is first energized and short-circuited after a period of about one minute, satisfactorily diminish thermal transients in the heaters. The sudden removal of the full voltage is not serious because the decrease in heater temperature is slow.

Thus it is seen that the main function of the motor-generator sets is to isolate the electronic equipment from severe voltage transients. Their elimination would result in a simpler, less expensive power system, requiring no maintenance of rotating machines for which electronic technicians have neither the skill nor the desire to learn.

The SAGE System for the air defense of the United States contains a large number of weapons-direction centers equipped with high-speed electronic digital computers for the continuous processing of vast quantities of information. It is necessary that substantially all voltage transients which might cause computer errors be prevented from entering the electronic system. At the first two centers, public utility power and local diesel generators operate in parallel, as shown in Fig. 1. Sufficient diesel capacity is connected to carry the load if utility power should fail. Following a thorough review of the load and reliability requirements, the design of the generating station for subsequent centers was modified and the utility tie was eliminated, as shown diagrammatically in Fig. 2. Identical 650-kw units generate at 480 volts to supply a peak load of 2,600 kw. The dual busses are normally closed to reduce motor-starting transients. The electrical arrangement of the motor-generator sets, interposed between the power sources and the electronic equipment, is presented in Fig. 3. The two computers are electrically separated from the input-output and display equipment which serves as the communication link between the computers, and people and other equipment within and without the site. The arrangement is the same for either generating scheme. The general distribution system for the 480-volt station is shown in Fig. 4. The philosophy of design and a description of the operation of the system was presented in an earlier article.[1]

With the elimination of the utility tie, voltage transients are produced only by disturbances within the site. It is possible that the occurrence rate of transients of sufficient magnitude to cause computer errors might be insignificant and that the motor-generator sets could be eliminated. The extensive investigation reported in this paper indicates that computer reliability will not depreciate if they are eliminated. Fig. 5 presents a proposed arrangement of the power system supplying the electronic equipment after the elimination of the motor-generator sets. Induction regulators maintain the filament heater voltage, and heater voltage increase is accomplished by impedance elements physically located in secondary distribution frames (not shown in the diagram). The dual busses are normally open to preserve a duplex system from the generating system down to the computers.

The investigation of computer reliability in a system without motor-generator sets was conducted in two parts. In the first, the effect of power transients on computer reliability was determined by obtaining the voltage tolerances of the computer and the transfer characteristics of the d-c supplies. In the second, calculations determined the physical limits in the distribution system within which voltage transients would be of sufficient magnitude and duration to cause possible computer error. The rate of occurrence of the transients was calculated from estimates of equipment fault rates. The combination of the results of the two parts gave the number of computer errors per year that would be produced.

Effect of Power Transients on Computer Reliability

With the elimination of the motor-generator sets, the inputs of the electronic power systems are subjected to all the fluctuations that occur on the generator busses. The fluctuations are continually produced by the starting and stopping of motors, by the switching off and on of all types of loads including those of the electronic equipment, and by faults in the equipment and in the distribution cabling. Most of these disturbances are of small magnitude and do not affect computer

---

Paper 57-676, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted October 24, 1956; made available for printing May 17, 1957.

J. J. GANO, formerly with Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass., is now with J. J. Gano and Associates, Cambridge, Mass.

S. T. Coffin of Lincoln Laboratory, Massachusetts Institute of Technology, investigated computer margins and transfer characteristics of d-c supplies. Jackson and Moreland, Inc., consultants to Lincoln Laboratory, calculated the voltage disturbances and estimated the fault probabilities. International Business Machines Corporation, as electronic equipment supplier, designed the distribution system to the electronic equipment without motor-generator sets.

The research in this document was supported jointly by the Army, Navy, and Air Force under contract with Massachusetts Institute of Technology.

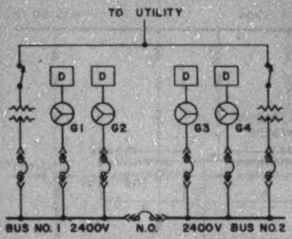

TO UTILITY
D
D
D
D
G1
G2
G3
G4
BUS NO. 1 2400V
N.O.
2400V BUS NO.2

Fig. 1. Arrangement of generating station having utility tie

Each generator=1,250 kw, 0.8 power factor
Each transformer=1,500 kva

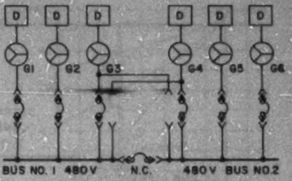

D
D
D
D
D
D
G1
G2
G3
G4
G5
G6
BUS NO. 1 480V
N.C.
480V BUS NO.2

Fig. 2. Arrangement of diesel generating station having no utility tie

Each generator=650 kw, 0.8 power factor

reliability. Portions of the electronic equipment are not subject to malfunction, even for the larger disturbances caused mainly by faults which the circuit breakers clear in 2, 6, or 14 cycles of 60-cycle power. Included among these are the cathodes of the thermionic tubes which have thermal time constants greater than 14 cycles. (The induction regulators, although they regulate the steady-state heater voltage, do not reduce the transients significantly.) Synchronous motors, which drive magnetic memory drums in the computer, have sufficient kinetic energy to override the severest transients. Transients on the input-output and display equipment create only momentary malfunctioning, the results of which are not serious and are immediately corrected. Therefore, this latter equipment, which is supplied from distribution frames *C* or *D*, was not included in the study. The effects were determined on but one computer because the computers are identical and independent in operation.

Computer D-C Margins

Since any transients which will pass through the d-c power supplies are slow as compared to the megacycle rate at which the computer circuits operate, a steady-state variation of a direct voltage will have roughly the same effect on the computer as a d-c power supply output transient of the same peak magnitude. Hence it is necessary only to vary all the direct voltages on a steady-state basis to reproduce the effect of a transient. This procedure gives a conservative result because a short-duration transient may cause only a transient computer error from which the calculating process can recover quickly. Because severe power transients reduce the direct voltage, margins for reduced voltage only were obtained The specifications of the d-c supplies required the same output voltage regulation for seven of the nine critical units. Of the remaining two, the specifications for one were twice as good, and for the other, half as good. For conservatism, they were all varied the same amount except the latter one which was varied twice as much.

Cognizance was taken of the fact that

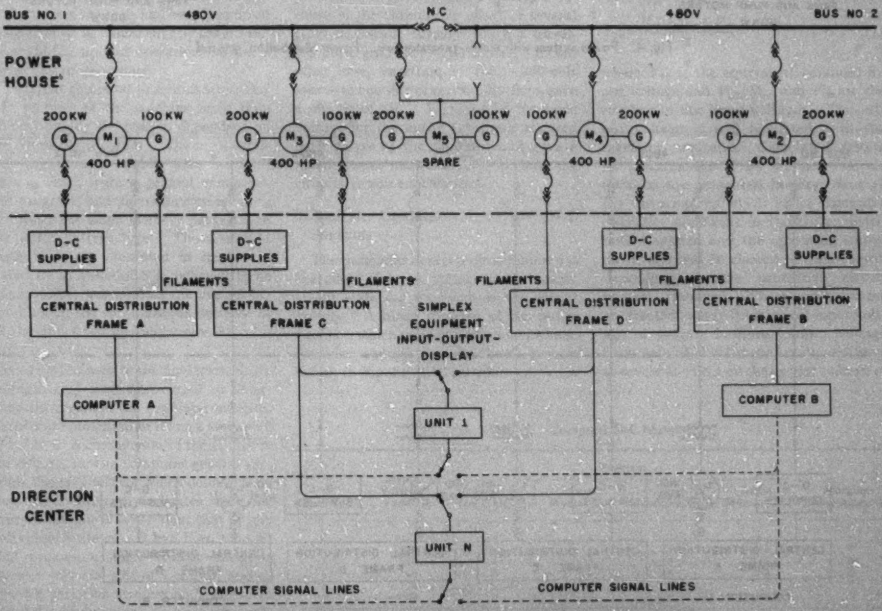

BUS NO. 1
480V
NC
480V
BUS NO 2
POWER HOUSE
200KW
100 KW
M_1
400 HP
200KW
100KW
M_3
400 HP
200KW
100KW
M_5
SPARE
100KW
200KW
M_4
400 HP
100KW
200KW
M_2
400 HP
G
D-C SUPPLIES
D-C SUPPLIES
D-C SUPPLIES
D-C SUPPLIES
FILAMENTS
FILAMENTS
FILAMENTS
FILAMENTS
CENTRAL DISTRIBUTION FRAME A
CENTRAL DISTRIBUTION FRAME C
SIMPLEX EQUIPMENT INPUT-OUTPUT-DISPLAY
CENTRAL DISTRIBUTION FRAME D
CENTRAL DISTRIBUTION FRAME B
COMPUTER A
UNIT 1
COMPUTER B
DIRECTION CENTER
UNIT N
COMPUTER SIGNAL LINES
COMPUTER SIGNAL LINES

Fig. 3. Power system with motor-generator sets. Computer power distribution

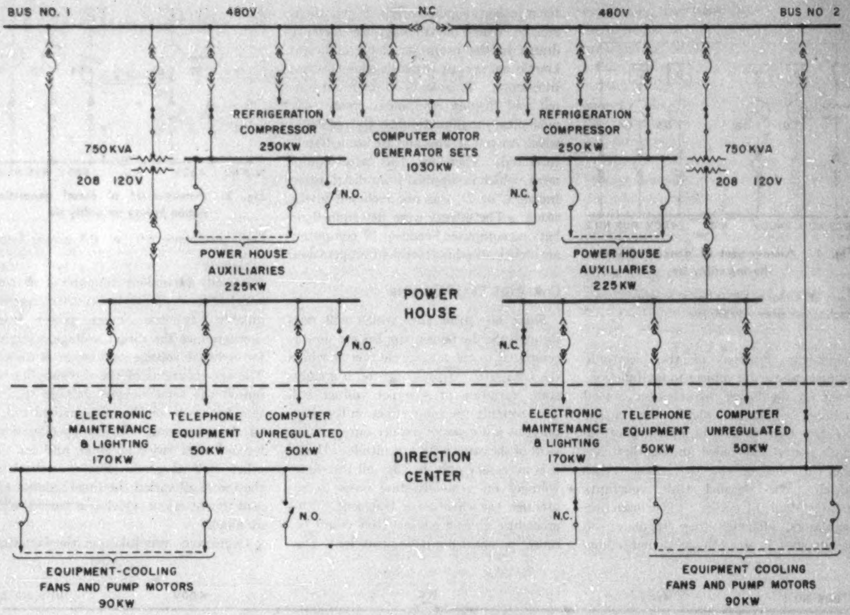

BUS NO. 1
480V
N.C
480V
BUS NO 2
REFRIGERATION COMPRESSOR 250KW
COMPUTER MOTOR GENERATOR SETS 1030KW
REFRIGERATION COMPRESSOR 250KW
750KVA
208 120V
750KVA
208 120V
N.O.
N.C.
POWER HOUSE AUXILIARIES 225KW
POWER HOUSE AUXILIARIES 225KW
POWER HOUSE
N.O.
N.C.
ELECTRONIC MAINTENANCE & LIGHTING 170KW
TELEPHONE EQUIPMENT 50KW
COMPUTER UNREGULATED 50KW
ELECTRONIC MAINTENANCE & LIGHTING 170KW
TELEPHONE EQUIPMENT 50KW
COMPUTER UNREGULATED 50KW
DIRECTION CENTER
N.O.
N.C.
EQUIPMENT-COOLING FANS AND PUMP MOTORS 90KW
EQUIPMENT COOLING FANS AND PUMP MOTORS 90KW

Fig. 4. Power system with motor-generator sets. Power distribution, general

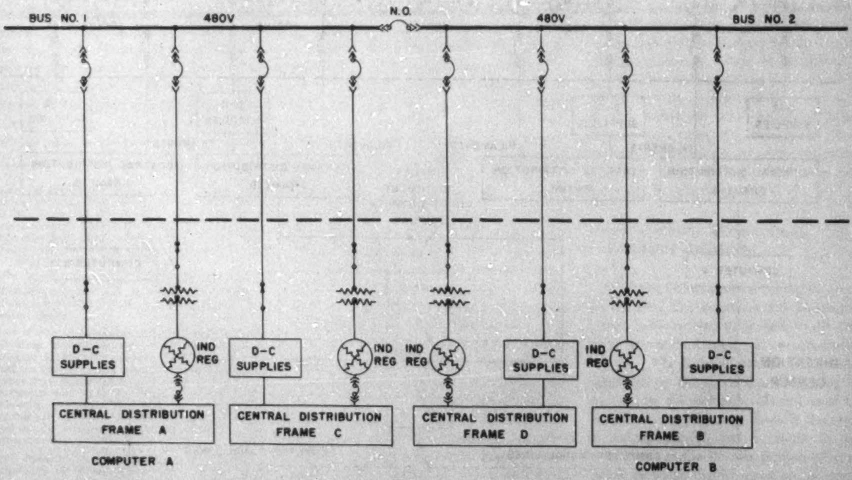

BUS NO. 1
480V
N.O.
480V
BUS NO. 2
D-C SUPPLIES
IND REG
D-C SUPPLIES
IND REG
IND REG
D-C SUPPLIES
IND REG
D-C SUPPLIES
CENTRAL DISTRIBUTION FRAME A
CENTRAL DISTRIBUTION FRAME C
CENTRAL DISTRIBUTION FRAME D
CENTRAL DISTRIBUTION FRAME B
COMPUTER A
COMPUTER B

Fig. 5. Power system without motor-generator sets. Computer power distribution

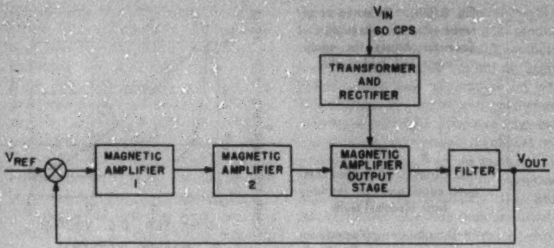


Fig. 6. D-c power supply block diagram

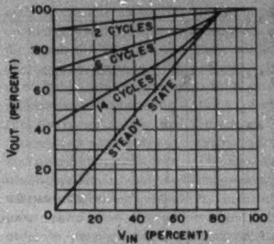


Fig. 7. D-c power supply transfer characteristics for disturbances of various durations

voltage margins are dependent on circuit design and the condition of the components. It was considered that any equipment which might be added or might replace other equipment would, at worst, have similar circuits with similar margins. Weak components are detected by a marginal-checking system which stresses a small section of circuitry by superimposing a variable direct voltage on the standard direct voltage. Replacements are made before margins are reduced appreciably. For purposes of collecting margin data, the circuit margins at the time of the tests had been permitted to deteriorate to values slightly lower than those during normal operation—again, a conservative procedure.

Different computer programs stress certain sections of the machine more than others, making the margin dependent on the operating program. The most sensitive programs available were selected, among which were a general computer-test program, and magnetic core memory and magnetic drum memory programs of the worst pattern type. The results of these tests are presented in item 1 of Table I. A value of 5%, believed to be conservative, was used as the margin beyond which the simultaneous decrease of all the direct voltages will cause the computer to fail.

As a preliminary to the final tests, single voltages only were first varied to determine the region of the computer margins. Results are presented in items 2 through 6 of Table I. A comparison of the figures in the column for the computer general program, the only test for which definite margins were obtained, indicates that the over-all margin is lower than that of any individual voltage. It had been thought that because a single voltage variation stresses only one side of a circuit, and an over-all variation stresses both sides, the over-all margin might be higher because the circuit remains balanced.

Another relation of interest is that between marginal checking data and single voltage variation, as shown in Table I in the columns giving computer general program and circuit marginal checking, items 4 and 6. The same program was used. The marginal-checking margins presented are for those circuits having the lowest margins of any of the computer circuits using the indicated voltage. Variation of the 90-volt source for a few gate circuits gave a substantial margin of 17%. Variation of the same voltage for the entire computer gave only an 8% margin because of the cumulative effect of several gates in cascade through which a traveling pulse gradually deteriorates. In another case, variation of the −300-volt source to one side of certain flip-flops gave a margin of 8%. Variation of the same voltage for the entire computer stressed both halves of the flip-flops and gave a much greater margin of 16% because the circuits remained balanced.

TRANSFER CHARACTERISTICS OF D-C SUPPLIES

The particular characteristic required is the effect of input voltage disturbances, steady-state and transient, on the output voltage. A block diagram of the power supplies, which use conventional magnetic amplifier circuits in a feedback loop, is shown in Fig. 6. (All rectifiers employ a 3-phase full-wave bridge.) Seven of the nine voltages have a total regulation specification of ±1.3% including ripple, slow load variations from no load to full, slow line variations of −7%, and step changes of load of 20% or step changes of line voltage of 5%.

The transfer characteristics of a d-c supply at full load are shown in Fig. 7. The curves are based on balanced input voltage for all three phases. For disturbances which produce unbalanced input voltages the following relation was used.

$$V_{in}=\frac{V_{ab}+V_{bc}+V_{ca}}{3} \qquad (1)$$

where V_{in} is the equivalent balanced input voltage and V_{ab}, V_{bc}, and V_{ca} are the rms line-to-line input voltages. The output voltage is well regulated until the regulator saturates. Calculated results for disturbances of 2-, 6-, and 14-cycle duration are presented because most of the disturbances which affect computer reliability are faults in the power distribution system and the operating equipment, and will be cleared in these times, depending on the protective circuit breaker involved. These disturbances are treated as step functions of input voltage of proper duration. Input voltage transients of short duration do not have as severe an effect on the output voltage of

Table I. Computer D-C Margins

Item No.	Direct Voltage Varied	Program				Marginal Checking, %
		Computer General, %	Core Memory, %	Core Memory, %	Drum Memory, %	
1	All	5–6	9	10	*	*
2	250	13	>13	*	8	*
3	150	16	>16	*	>16	*
4	90	8	>8	*	>8	17
5	−150	8	>8	*	7	*
6	−300	16	>16	*	8	8

* Margin not taken.
> Margin greater than.

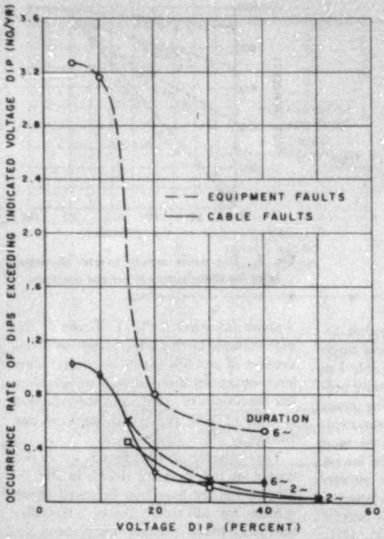


Fig. 8 (left). Estimated occurrence rates of voltage dips at generator busses for equipment and cable faults

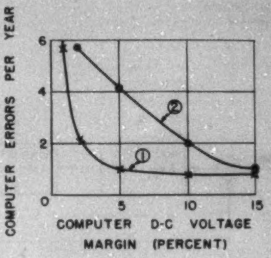


Fig. 9 (right). Estimated computer errors per year resulting from electrical faults

1—D-c power supply performance as given in Fig. 7
2—D-c power supply performance half as good

a d-c supply because the large output capacitance prevents the voltage from dropping appreciably before the input voltage has recovered. Extensive tests were conducted to verify the calculations.

Voltage Disturbances on the Generator Bus

To estimate the number of computer errors per year caused by disturbances in the power system, a prediction of the frequency of disturbances having sufficient magnitude and duration to penetrate the d-c supplies is necessary. Results presented are based on the operation of two generators on a bus. Calculations on disturbances created by the synchronization of diesel generators, starting of large motors, transfer of large loads from one generator bus to the other and inrush currents to transformers indicate that voltage drops will not exceed 8% and should offer no trouble. The problem is then reduced to that of predicting the number of faults in the power distribution system and auxiliary equipment that will dip the voltage on the generator bus sufficiently.

VOLTAGE DROPS DURING SYSTEM FAULTS

All types of faults were investigated, using the method of symmetrical components in the calculations. The value of resistance selected for the diesel-generator grounding scheme was sufficiently large to limit fault currents, but sufficiently small to assure fast tripping of the circuit breakers in the feeder cables during line-to-ground fault conditions; at the same time, the value is in the range where the advantages of grounding in minimizing the effect of line-to-ground transient voltages is obtained.

For each type and duration of fault, the equipment and length of cable was determined within which the occurrence of a fault will dip the voltage beyond computer tolerances.

FAULT OCCURRENCE RATES

There remains the determination of the rates of fault occurrences and the application of these rates to the amount of equipment and cable exposed to arrive at the frequency of computer errors due to electrical faults. Because there are very few published statistics on fault rates, the estimates of a group of qualified engineers with long experience in the operating field were obtained. The averages of the fault rates for cables and various equipment are presented in Table II. In general, the estimates lean toward the conservative side. About 80% will be line-to-ground faults and the rest will be nearly equally distributed among line-to-line, 3-phase, and double line-to-ground faults.

Estimates of occurrence rates of voltage drops that will exceed a prescribed value for cable or equipment are graphically summarized in Fig. 8. The path of the curves between points is uncertain. A majority of the critical voltage drops are the result of faults in equipment supplied by circuit breakers set for 6 or 14 cycles which are combined into one curve. However, the occurrence rate decreases rapidly for drops exceeding 10 to 20% of rated bus voltage because there is very little equipment and cable sufficiently close to the generator busses to produce severe drops. The small number for overall occurrence rates results because a large percentage of the faults are line-to-ground.

Estimate of Computer Errors

An estimate of the number of computer errors per year produced by power transients in the system without motor-generator sets for this installation can be readily obtained from the curves of Figs. 7 and 8. The results for a wide range of d-c margins on the computer are presented in curve *1* of Fig. 9. For the established margin of 5%, only one computer error per year can be expected. Since the power system will operate with separated generator busses and calculations were based on the equipment for the entire station, each computer will err but once every 2 years. Even if the margin is reduced to 2%, the number of errors, though doubled, is still low. A reduction of margin below this point would be pre-

Table II. Estimated Probabilities of Fault Occurrences in Equipment as Used in Studies of Computer Errors

Equipment Involved	Estimated Fault Occurrence
Cable in iron conduit...	0.2 per 1,000 circuit-feet-year
Motors, including terminations...	0.05 per motor-year
Circuit breakers, including terminations...	0.01 per breaker-year
Transformers..........	0.03 per transformer-year

carious because of the rapid rise of the curve. Curve *1* of Fig. 9 also indicates that there is very little to be gained by increasing the computer margin above 5%. The co-ordination between the performances of the computer and the d-c supplies appears good for the designed power generation and distribution system. Curve *2* gives an estimate of computer errors if the d-c supplies had half as good regulation as that described in Fig. 7. The number of computer errors per year remains low. A further deterioration in the regulation of the supplies will produce unreliable results because disturbances other than those due to faults should be considered. These estimates have been calculated on the basis of two diesel generators connected to each of the generator busses. With three generators the values are reduced several per cent.

Conclusions

Using conservative methods throughout the investigation, it was concluded that the elimination of motor-generator sets as voltage transients buffers will produce one error every 2 years for each computer in SAGE installations having no utility power source and generation on site. Motor-generator sets and associated switchgear with their extensive relay control systems and voltage regulators will probably produce about the same number of computer errors because of equipment malfunction. Furthermore, the estimated number of errors thus produced is far below that which may be expected as a result of component difficulties in the vast electronic system. The decision to eliminate the motor-generator sets is logical. The result is a simpler power system with reduced initial costs, maintenance, and power consumption, and a probable increase in reliability because of operation with separated generator busses.

Most commercial installations use a public utility service as the source of power. Thus the number of computer errors is dependent on the rate of occurrence of voltage disturbances within and without the site as well as on computer d-c margins and transfer characteristics of d-c supplies. The results of these studies offer a guide for some of these items. In most public utility systems, a number of interconnected generating stations constitute the power source, and the energy is transmitted over extensive transmission and distribution networks. The lines are usually well maintained and designed with adequate lightning protection; in urban areas, they often run underground. In order to pass specific judgment on the probability, magnitude, and duration of voltage dips and outages in a particular location, however, some details concerning the utility system and the vulnerability of the area to lightning and sleet storms, and other weather conditions must be available.

Reference

1. POWER GENERATION AND DISTRIBUTION AT AIR-DEFENSE COMPUTING CENTERS, J. J. Gano. *Electrical Engineering*, vol. 75, Dec. 1956, pp. 1098–1102.

The Canadian Transcontinental Microwave System

S. BONNEVILLE
NONMEMBER AIEE

IN a paper[1] reviewing the subject of transcontinental telephone communications in Canada reference was made to the long-term relief plans of the Trans-Canada Telephone System. It was pointed out that this envisaged a coast-to-coast radio-relay system carrying television and *L*-carrier facilities. The construction of this system is now well advanced and is scheduled for completion in 1958.

It is the purpose of this paper to review the features which have been adopted to meet the particular conditions. Information concerning the basic *TD-2* system has already been published.[2–4] H. E. Curtis and others have expanded on some of the problems encountered in the stacking of television and message facilities. Spur route arrangements associated with the microwave system to provide telephone and television service along the route at points which are remote from the terminals have been discussed by Dinnin.[5]

Operating Arrangements

Telephone communication in Canada is provided by seven major telephone systems extending from the Pacific coast to the Atlantic coast. The British Columbia Telephone Company operates on the west coast within British Columbia. In each of the three prairie Provinces of Alberta, Saskatchewan, and Manitoba, the systems are government operated. The Bell Telephone Company of Canada operates in Ontario and Quebec. Service in the Maritimes is given by The New Brunswick Telephone Company Ltd., in New Brunswick, and the Maritime Telegraph and Telephone Company which operates in Nova Scotia and Prince Edward Island.

In 1932 the Trans-Canada Telephone System was formed in order to provide and maintain satisfactory long-distance facilities across Canada, and to establish uniform operating procedures which would speed the handling of long-distance traffic. Although each administration provides the plant and equipment required to furnish service in its own territory, the construction, operation, and maintenance of the Canadian transcontinental microwave system is being undertaken by the Trans-Canada Telephone System.

SYSTEM LAYOUT

Fig. 1 shows the route which has been selected for the system. Cities shown by circles are drop-off points for either telephone or television. The total system mileage, including spurs, will be about 4,800 miles. There will be approximately 137 repeater points on the main route. The Buffalo-Toronto - Ottawa Montreal section was placed in service in 1953. This was extended to Quebec City in 1954. The Toronto-Winnipeg section was placed in service for television in September 1956, and for message service in November 1956. If the present construction schedule is maintained, the entire system will be completed in 1958.

Paper 57-706, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 20, 1957; made available for printing April 25, 1957.

S. BONNEVILLE is with the Bell Telephone Company of Canada, Montreal, Que., Canada.

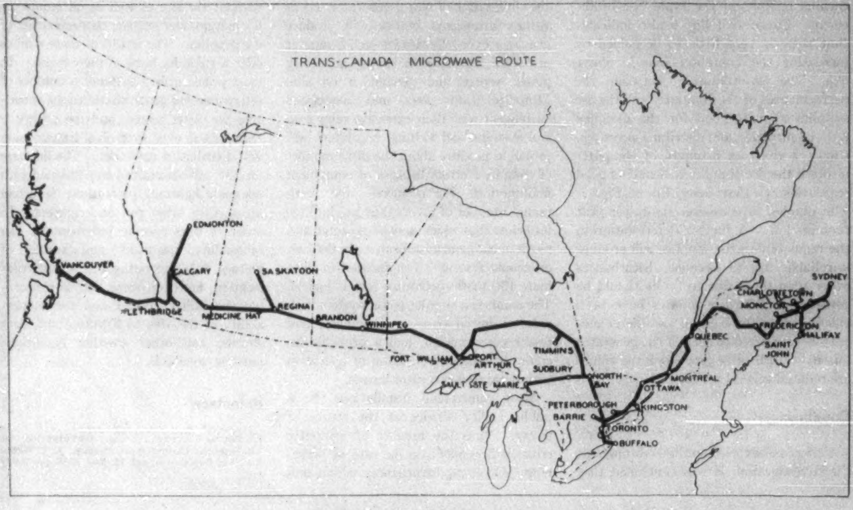

TRANS-CANADA MICROWAVE ROUTE
EDMONTON
VANCOUVER
CALGARY
SASKATOON
LETHBRIDGE
MEDICINE HAT
REGINA
BRANDON
WINNIPEG
FORT WILLIAM
PORT ARTHUR
TIMMINS
SUDBURY
SAULT STE MARIE
NORTH BAY
PETERBOROUGH
BARRIE
TORONTO
BUFFALO
KINGSTON
OTTAWA
MONTREAL
QUEBEC
SAINT JOHN
FREDERICTON
MONCTON
CHARLOTTETOWN
HALIFAX
SYDNEY

Fig. 1. Trans-Canada microwave route

Microwave Equipment

To meet acceptable objectives for distances up to 4,000 miles and at the same time provide large blocks of telephone circuits and several television circuits, it was decided to use Northern Electric *TD-2* microwave equipment. This is a high-capacity system designed by the Bell Telephone Laboratories to meet high over-all standards. This equipment operates in the common-carrier 4,000-mc (megacycle) band as shown in Fig. 2 As this figure shows, the system has a capacity of six radio channels in each direction. Each 2-way channel handles two television programs or 600 telephone circuits. One 2-way channel of the six is normally reserved for protection. Thus a fully loaded system has five working channels with a capacity of 3,000 telephone circuits or five 2-way television channels, or a combination of both. However, as each channel has an effective base bandwidth of approximately 10 mc and as a television program only occupies a 4-mc bandwidth, it is possible under certain conditions to carry a television program and a moderate number of telephone circuits on the same channel simultaneously, the controlling factor being the noise appearing in the telephone circuits. This process is known as "double decking" and will be employed initially on several sections of the transcontinental system.

Fig. 3 shows a typical repeater station in the Toronto-Winnipeg system. Note the large horn reflector antenna. This antenna has an effective gain of 39.5 db (decibels) referred to an isotropic radiator.

Fig. 4 shows a *TD-2* equipment bay. Fig. 5 is a block diagram of the radio system; note that the signal is not brought down to base band, but is passed through at an intermediate frequency of 70 mc. The transmitter output power is in the order of 1/2 watt.

In 1950, during the initial planning of the Buffalo-Toronto-Ottawa-Montreal system, the problem of selecting radio-relay equipment which would most readily meet the requirements was encountered. At that time, the tremendous demands for television and telephone circuits which were to materialize in the next few years could not be foreseen. Therefore some of the reasons for the selection of *TD-2* equipment for this first radio-relay system are given.

It was found that the *TD-2* system, by using regular and slot frequencies (see Fig. 6) made a more efficient use of the spectrum than other systems studied at that time. Also, the *TD-2*

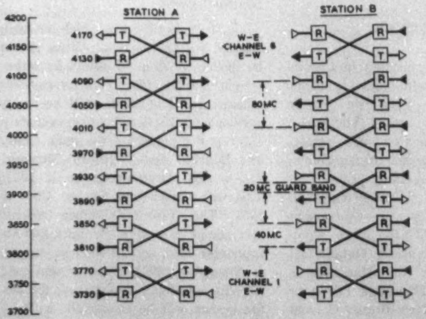

STATION A
STATION B
4200
4150
4100
4050
4000
3950
3900
3850
3800
3750
3700
4170
4130
4090
4050
4010
3970
3930
3890
3850
3810
3770
3730
T
R
W-E CHANNEL 6 E-W
80 MC
20 MC GUARD BAND
40 MC
W-E CHANNEL 1 E-W

Fig. 2. TD-2 radio-frequency plan

Fig. 3 (left). Repeater station, Toronto to Winnipeg

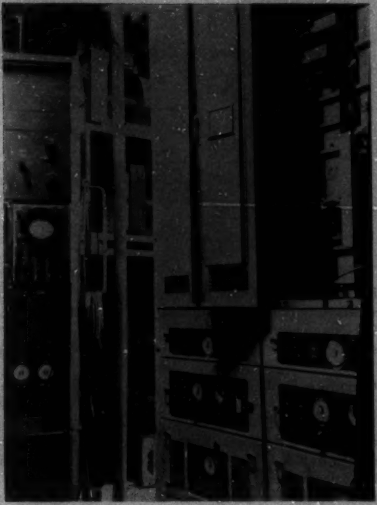


Fig. 4 (right). TD-2 equipment bay

system, by employing one protection channel for five working channels, assured a high usage of the channel capacity.

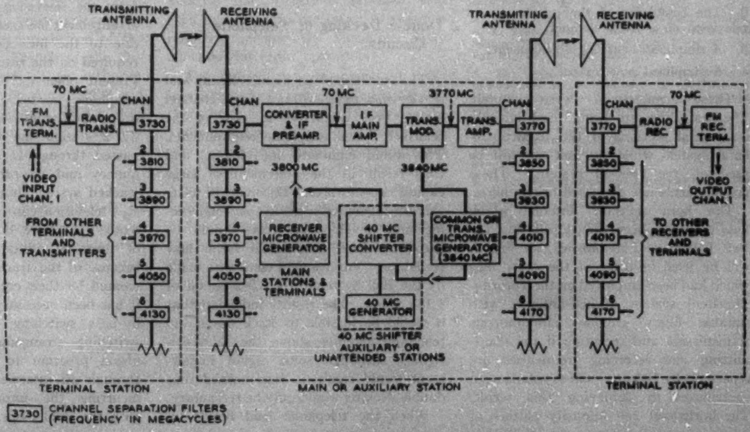
Delay-lens antenna had a much greater directivity pattern and higher back-to-front ratio than parabolic antennas in common use for microwave systems. This high front-to-back ratio is important in preventing interference at repeater stations.

The linearity of the *TD-2* system was found to be better than most of the other equipment studied. This, of course, is one of the reasons for the greater telephone capacity of the *TD-2* equipment as compared with other systems.

At the time the microwave applications were studied in Canada, a fairly long *TD-2* system was already in operation in the United States between Boston, New York, and Chicago. Operating results showed that satisfactory transmission was being obtained over this distance. Further, the equipment was continually being improved as operating experience became available. Plans were already under way to extend this system to the west coast. Many of the other systems were untried and would likely require considerable engineering effort before satisfactory service could be obtained even on a relatively short route.

These are a few of the factors which influenced the choice in selecting the *TD-2* equipment for the Buffalo-Toronto-Ottawa-Montreal system. Actual operat-

Fig. 5. Radio system block diagram


TRANSMITTING ANTENNA
RECEIVING ANTENNA
TRANSMITTING ANTENNA
RECEIVING ANTENNA
70 MC
FM TRANS. TERM.
RADIO TRANS.
CHAN 1
3730
2
3810
3
3890
4
3970
5
4050
6
4130
VIDEO INPUT CHAN. 1
FROM OTHER TERMINALS AND TRANSMITTERS
TERMINAL STATION
CHAN 1
3730
CONVERTER & IF PREAMP.
70 MC
I F MAIN AMP.
TRANS. MOD.
3770 MC
TRANS. AMP.
CHAN 1
3770
2
3810
3
3890
4
3970
5
4050
6
4130
3800 MC
3840 MC
RECEIVER MICROWAVE GENERATOR
MAIN STATIONS & TERMINALS
40 MC SHIFTER CONVERTER
40 MC GENERATOR
40 MC SHIFTER AUXILIARY OR UNATTENDED STATIONS
COMMON OR TRANS. MICROWAVE GENERATOR (3840 MC)
2
3850
3
3930
4
4010
5
4090
6
4170
MAIN OR AUXILIARY STATION
CHAN 1
3770
RADIO REC.
70 MC
FM REC. TERM.
VIDEO OUTPUT CHAN. 1
2
3850
3
3930
4
4010
5
4090
6
4170
TO OTHER RECEIVERS AND TERMINALS
TERMINAL STATION
3730 CHANNEL SEPARATION FILTERS (FREQUENCY IN MEGACYCLES)

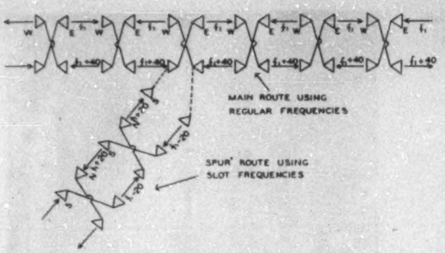

W
E
f1
f1+40
f1-20
f1+20
S
MAIN ROUTE USING REGULAR FREQUENCIES
SPUR' ROUTE USING SLOT FREQUENCIES

Fig. 6 (left). Regular- and slot-frequency arrangement

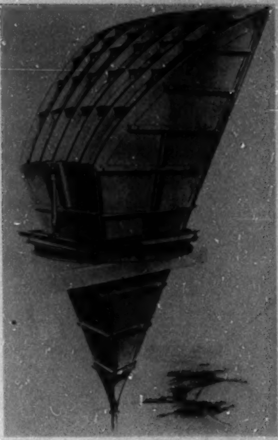


Fig. 7 (right). Horn reflector antenna

ing experience on this system has been satisfactory and has resulted in adopting *TD-2* equipment for the entire continental system.

Canadian Applications

Studies made during the initial planning stages of the transcontinental system disclosed that originally there would be a requirement for the provision of transmission facilities for a television program and a limited number of telephone circuits. However, the telephone circuit requirements were expected to grow, and with the adoption of nationwide customer toll dialing a large block of circuits would be required. When the great distance involved was considered, it soon became apparent that some simplification of the existing *TD-2* system would be required if the first costs of the system were to be kept to a reasonable figure. Several of the following possibilities were apparent:

1. The use of a single antenna for transmitting and receiving.
2. The combining of the telephone and television on a single channel.
3. A simplified switching arrangement.
4. A simplified power plant.

These arrangements were adopted. The first three are being used initially on several sections of the transcontinental system while the last one will be used on the over-all system. These arrangements are discussed briefly here.

The horn reflector antenna which has recently been developed by the Bell Telephone Laboratories, see Fig. 7, will be used in place of the delay lens which had been employed on the Toronto-Montreal system. This antenna with suitable filters provides simultaneous transmission and reception if the transmitting and receiving frequencies are separated by 100 mc. Cross polarization contributes in achieving this result. The horizontal field-intensity pattern of the horn reflector antenna is shown in Fig. 8.

Single-Antenna Technique

The requirement that the transmitting and receiving frequencies be separated by 100 mc when using a single antenna for transmitting and receiving has necessitated using a different frequency arrangement from that previously shown. Fig. 9 shows the frequency arrangement which has been adopted. Note that the capacity of the system has been cut from six to three 2-way channels. This is not as serious as it seems because the system capacity can be increased to six channels at any time by installing a second horn reflector antenna.

The use of a single antenna for both transmitting and receiving, therefore, provides a considerable saving in the initial capital expenditure. Only one antenna is required at terminals and two at repeater stations. The transmission characteristics of the system will not be affected by employing this technique.

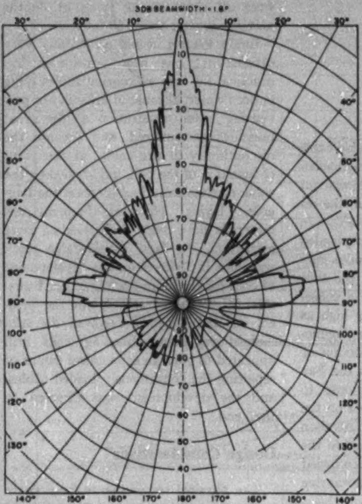
Double Decking of Telephone Circuits

The stacking of telephone circuits and a television program on a single channel looked very attractive in view of the relatively small telephone requirements. This would eliminate one channel and thus result in the deferment of large capital expenditures. However, in using this arrangement considerable engineering and development costs would be encountered as this would be a new technique. Various field tests were made by the Bell System on a *TD-2* circuit of 1,133 miles. The results indicated that it would be feasible to stack up to 120 telephone circuits above the television signal. Although some signal impairment would result, it was realized that double decking would only be temporary.

When the telephone load builds up to 120 circuits an additional channel is required and the system then reverts to the standard arrangement of separate channels for both television and message. From present circuit growth it is expected that most of the stacking will be eliminated by 1958 or 1959. In view of the savings factor, it was decided to proceed with this technique using as many existing standard components as possible in order to keep development costs to a minimum.

Separation of Telephone and Television

The separation of the telephone circuits from the television signal in a stacked arrangement has created several problems due to the number of drop-off points required on the transcontinental system. Present plans call for 12 points where telephone or television can be either dropped off or introduced. On a standard *TD-2* system the through circuits are passed through at an intermediate frequency and no problem exists. On a stacked system, however, the telephone and television can only be split at base band. Therefore, because of the number of drop-off points, or more precisely because of the transmission impairment created by their connecting video loops, it has been necessary to develop a video by-pass switching arrangement. This switching arrangement enables a television program to be passed directly through an intermediate point without incurring the impairments associated with the local loops.



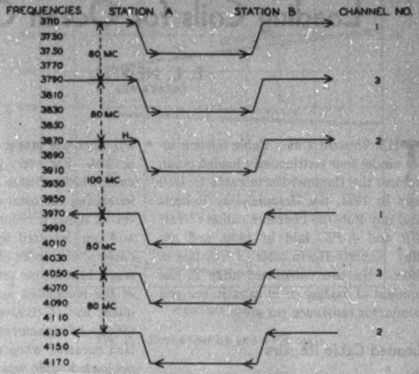


Fig. 8 (left). Radiation pattern of KS-15676 horn reflector antenna; horizontal directivity and polarization; 3,740mc

Fig. 9 (above). Frequency assignment, single horn antenna

Protection and Automatic Switching

As only two channels, an operating and a protection channel, will be provided initially, it is possible to use a simplified form of automatic switching. This switching system will form part of the more complicated and expensive unit which will be required when the protection channel is called upon to protect more than one working channel.

Licensing

The licensing of radio systems in Canada is under the jurisdiction of the Telecommunications Branch of the Department of Transport. Their authority is covered in the Radio Act of 1938 and the regulations thereunder. This subject is reviewed in more detail by Brant.[6]

Radar Interference

The growing use of radar for defense, aeronautical and meteorological purposes has necessitated serious consideration being given to possible interference to the microwave systems. Although the band in which these systems operate is outside that used by the telephone common-carriers, side-mode and harmonic radiation results in interference in the common-carrier bands.

The National Research Council in Ottawa and the Bell Telephone Laboratories in the United States have carried out extensive tests to determine the extent of the interference problem. These tests, although not completed, indicate that a serious problem may exist unless means are taken to suppress radiation outside the frequency band allocated for radar operation.

Although no definite reports of radar interference have been received in Canada there have been several instances in the United States. In one case interference was experienced from radar equipment some 60 miles distant. In some cases it has been possible to eliminate this interference by changing the frequency of the radar or the telephone equipment. On fully loaded telephone routes, changing frequencies is not a solution to the radar problem. The answer would appear to be to correct the trouble at the source, i.e., to eliminate or suppress the undesired emissions. The Bell Telephone Laboratories has recently developed an evacuated filter for one type of radar that reduces the interfering spurious products by some 70 to 80 db.

References

1. DEVELOPMENT OF TRANSCONTINENTAL COMMUNICATIONS IN CANADA, J. W. Noyes, G. Gaudet, S. Bonneville. *AIEE Transactions*, vol. 75, pt. I, July 1956, pp. 342–52.

2. THE TD-2 MICROWAVE RADIO RELAY SYSTEM, A. A. Roetken, K. D. Smith, R. W. Friis. *Bell System Technical Journal*, New York, N. Y., vol. 30, Oct. 1951, pp. 1041–77.

3. THE TRANS-CANADA TELEPHONE SYSTEM'S MICROWAVE RADIO PROJECT, A. J. Groleau. *The Engineering Journal*, Montreal, Que., Canada, vol. 39, Aug. 1956, pp. 1001–09, 1940.

4. A BROAD-BAND TRANSCONTINENTAL RADIO RELAY SYSTEM, T. J. Grieser, A. C. Peterson. *Electrical Engineering*, vol. 70, Sept. 1951, pp. 810–15.

5. LIGHT-ROUTE MICROWAVE SYSTEMS IN CANADA, A. J. Dinnin. *AIEE Transactions*, vol. 76, pt. I. 1957 (*Paper no. 57-707*).

6. A CANADIAN POINT OF VIEW ON RADIO-FREQUENCY SPECTRUM MANAGEMENT, C. M. Brant. *Ibid.*, Sept. 1957, pp. 455–61.

Loading Coils for Ocean Cables

E. L. NEWELL
MEMBER AIEE

THE Western Union Cable System includes four continuously loaded ocean cables: the Hammel-Horta cable (*1 HO*) laid in 1924, the Hammel–Bay Roberts and Bay Roberts–Penzance cables (*2HM-BR* and *4 PZ*) laid in 1926, and the Bay Roberts–Horta cable (*2 HO*) laid in 1928. No two cables are alike in the amount of loading or in capacitance and conductor resistance per mile.

Loaded Cable Repairs

Enough cable normally remains after the initial laying to provide adequate stock for repairs. Sufficient heavily armored loaded cable of each type is in storage at the cable depots for all anticipated repairs in the shore ends of the loaded cables. The stock of loaded cable suitable for deep-sea repairs, however, was almost entirely depleted in repairing extensive damages suffered in the November 1929 submarine earthquake off the North American continental shelf. Since then, repairs in deep-sea sections of the loaded cables have been made with nonloaded cable.

Aside from the earthquake incident, there have been only four interruptions in the deep-sea sections of the loaded cables, two in 1951 and two in 1956, with a total replacement of nearly 50 miles of cable. A gradual increase in the number and extent of the repair requirements may be expected, because of the accumulation of aging effects.

In the deep-sea sections of the cable, that is, at depths greater than 1,000 fathoms (6,000 feet), each cable repair usually adds at least 5 miles of nonloaded cable. This added length is twice the depth of the water plus the amount of cable that may have to be picked up before a fault is located. Occasionally up to 30 or more miles of cable may have to be replaced because of mechanical damage or deterioration from age. Repairs in shallow water near shore usually involve much less than 5 miles of cable.

Nonloaded stock repair cable can be used for the repair of loaded cables, but only at the expense of signal deterioration. For every 5 miles of nonloaded cable inserted in the loaded cable, which increases the total cable length by that amount, the received signal amplitude is reduced by 4 or 5%, as compared to a reduction of 2% when the proper type of loaded cable is used. For every 5 miles of nonloaded cable which replaces loaded cable without increasing the total cable length, the received signal is reduced by 2 to 3%. In addition, inserted lengths of nonloaded cable of the order of 5 to 30 miles may result in multiple reflections at the ends of the nonloaded section, with a consequent loss of received signal as high as 50%. An instance of a 50% signal reduction occurred when a 22-mile section of nonloaded cable was inserted in the Bay Roberts–Penzance 1926 cable in order to reroute it around the trawler area on the Newfoundland fishing banks. Less than half of this loss can be attributed to the attenuation of the added nonloaded cable; the major part of the loss was due to the multiple reflections.

While it is undesirable from a transmission point of view to use nonloaded cable for loaded cable repairs, the cost economy is considerable. Nonloaded repair cable costs about $3,000 per nautical mile; continuously loaded cable normally about $4,500. However, the present cost of the loaded cable would be considerably more, because special machinery for forming, applying, and annealing the loading material would first need to be built. To this initial cost of the loaded repair cable must be added carrying charges on an adequate stock, and the expense of storing it at the cable depot and on the cable ship.

Lump Loading

Nonloaded repair cable must be carried in stock in considerable amount for use in the older nonloaded cables. Of course, this same stock would be adequate also for the repair of the loaded cables if suitable loading coils were inserted in the nonloaded repair cable to make it electrically equivalent to the continuously loaded cable. Such lump-loading has been widely used in telephone land lines, and to some extent in submarine cables.

Loading coils were first used in a submarine cable in 1906 when Siemens and Halske in Germany laid a lump-loaded telephone cable 9.3 miles long across Lake Constance. Subsequently several lump-loaded telephone cables were laid across the English Channel and across the Irish Sea by the British Post Office. These were all short cables lying at depths nowhere greater than 150 fathoms.

Lump loading, while electrically very satisfactory, has been considered impractical for long submarine cables lying in deep water, because of difficulty of preventing water penetration at the great pressures encountered, and because the discontinuity in the cable, at the point of insertion of the coil, might weaken the cable mechanically and increase the hazards of laying. However, the many new materials have made lump loading more attractive. In fact, considerable work was done in 1946–47 on the development of coils for the lump-loading of our transatlantic nonloaded cables, but the project was abandoned in favor of the submerged repeaters. Loading coils still appeared to offer an economical means of repairing the continuously loaded cables, and the development was directed to that end.

Design Considerations

The physical requirements for a deep-sea loading coil are severe. The coil must withstand hydrostatic pressures of 3 or 4 tons per square inch without water leakage, or damage to the stress-sensitive high-permeability alloys which must be used as core material in the coil in order to obtain high electrical efficiency with small size. The coil should not be permanently affected by the great stresses, either longitudinal or bending, that may occur during the laying of the cable in deep water. Also, the coil should not introduce into the cable any mechanical discontinuity which might interfere with the normal adjustment of the core and armor to the decreasing tension in the cable as it passes from the cable ship to the ocean floor. To do so might cause a serious kink in the cable. It is very desirable, too, that the loading coils require no special handling as the cable is being laid.

Materials used in construction of the coils must not deteriorate with age, at

---

Paper 57-705, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 6, 1957; made available for printing April 26, 1957.

E. L. NEWELL is with the Western Union Telegraph Company, New York, N. Y.

The author acknowledges the contributions of many persons, in their special fields, to the loading coil project, and especially mentions G. A. Randall and P. H. Wells, for their calculations to determine the required constants of the coil; R. C. Taylor for the design of the coil winding and core, and T. Rystedt who was responsible for many of the precision construction features of the coils. The method of jacketing and armoring, and of anchoring the cable stubs in the end plugs, was developed jointly with the Simplex Wire and Cable Company, Cambridge, Mass.

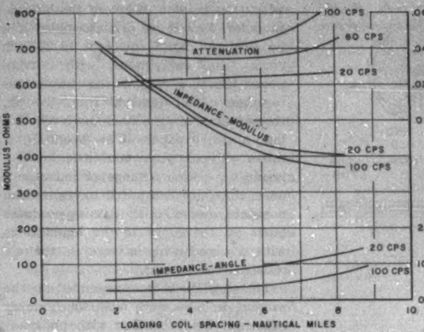


Fig. 1. Characteristics of lump-loaded stock repair cable

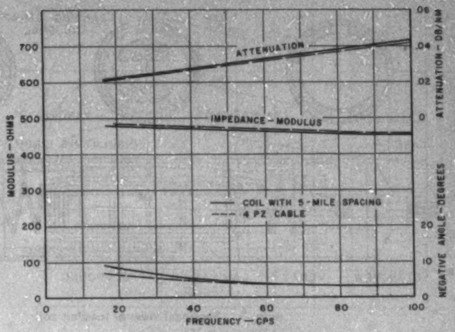


Fig. 3. 5-mile spacing and 4 PZ cable

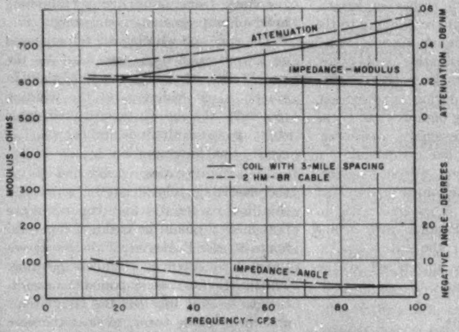


Fig. 2. 3-mile spacing and 2 HM-BR cable

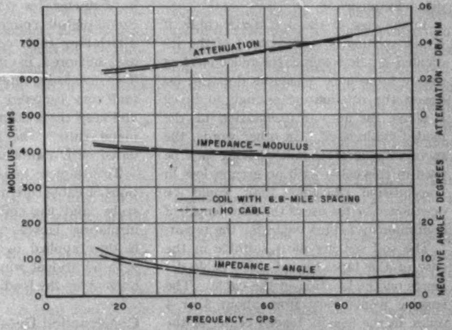


Fig. 4. 6.8-mile spacing and 1 HO cable

least not more rapidly than the cable itself, and must be free of any undesirable chemical interactions, either during construction or after laying.

The most practical method of meeting these conditions appeared to be to make the loading coil an integral part of the cable with the armor wires extending, continuous and unbroken, over the coil. At the bulge created by the coil, additional wires would be interposed in the armor to provide complete coverage. The coils could be spliced into the nonloaded stock repair cable as required, and then be laid as part of the cable, without special attention.

The typical nonloaded cable used for repairs is 1 inch in diameter over-all. An enlargement to 3 inches at the point of insertion of the loading coil is about the maximum bulge that can safely pass through the cable ship gear during laying of the cable. More particularly, the diameter over the armor was limited to something less than 2.75 inches, since that was the maximum that could be handled in the available armoring machines. Then, allowing space for armor wires, padding, and insulating jacket, the diameter of the loading coil itself could not exceed 1.5 inches. Some sort of protective case being assumed, the diameter of the core and winding would be limited to about 1.25 inches. The length of the coil would depend on the amount of core material and winding which was required to give the desired inductance and resistance.

The electrical characteristics of the loading coil should be such that, when the coil is inserted in nonloaded stock repair cable, it makes the repair cable equivalent to the continuously loaded cable in impedance, and not appreciably greater in attenuation. Calculations showed that coils inserted in the repair cable at 5-mile intervals should have an inductance of 0.3 h (henry) and a resistance of not more than 7 ohms, a-c plus d-c, at 60 cps (cycles per second), and 1 to 20 ma (milliamperes), in order to match the *4 PZ* loaded cable. The same coil inserted with 3-mile spacing would match the *2 HM-BR* cable, and with 6.8-mile spacing the *1 HO* cable. These spacings would be sufficient to avoid reflection losses, which increase attenuation or affect waveshape; for all practical purposes these spacings would give the effect of a smooth (continuously loaded) cable within the range of telegraph frequencies.

Fig. 1 shows how the impedance and attenuation of the lump-loaded repair cable would vary with the spacing of 0.3-h 7-ohm loading coils. Good impedance and attenuation matches can be obtained with the loading coils at 3-, 5-, and 6.8-mile spacing for the *2 HM-BR*, *4 PZ*, and *1 HO* cables respectively. Figs. 2, 3, and 4 in-

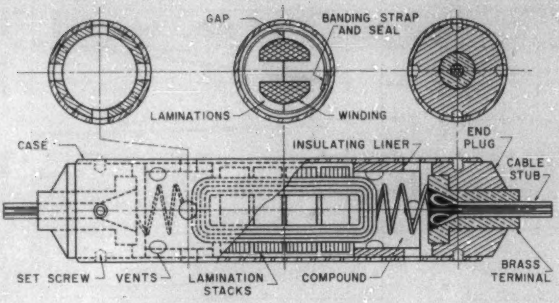


Fig. 5. Sectional view of loading coil

dicate the degree of equivalence that would be obtained over the range of telegraph frequencies.

In the case of the *2 HM-BR* cable if 5-mile spacing is used instead of the indicated 3-mile spacing, the reflection loss due to impedance mismatch does not increase the attenuation enough to bring it above the normal attenuation of the loaded cable itself. In other words, the more economical 5-mile spacing can be used in this cable with no greater loss in transmission efficiency than would be occasioned by the use of the same length of continuously loaded cable for the repair.

The coil spacing most suitable in the case of the taper-loaded *2 HO* cable depends on the location in the cable. The spacing would vary from about 3 or 4 miles, in the heavily loaded center of the cable, to 8 or 10 miles in the more lightly loaded end sections.

The Loading Coil

Figs. 5 and 6 show the loading coil (less jacket and armor) that was developed to meet the afore-mentioned conditions. The winding, with its laminated core of high-permeability steel, is located midway in a stainless-steel tube 1.5 inches in diameter by 6.25 inches in length. Spiral pigtails connect the winding to the 4-inch-long cable conductor stubs which are anchored in insulating end plugs. Two insulating liners keep the winding and core centered during construction, and keep the winding pigtails clear of the metal tube. The assembly is impregnated with a rubber-like compound.

To complete the coil, 45-foot-long sections of cable core are joined to the short stubs, and the coil is covered with an insulating jacket. Regular wire armor is then applied over the entire length, with additional wires interposed at the bulge over the loading coil.

Construction Details

Some of the winding and core construction details may be seen in Fig. 7. At the lower left is the winding form, and the clamp which holds the winding in shape during its initial impregnation.

The core is made up of 0.007-inch-thick modified *E*-type laminations of annealed mu-metal or molybdenum permalloy. In building the core the laminations for one coil are first assembled in a jig (see top Fig. 7) in eight 15/32-inch stacks, and impregnated with a compound that firmly binds the laminations together. After removal from the jig, the gap faces of each stack are leveled on fine, abrasive paper. This operation is rather critical, since too little leveling results in a large change of inductance when the coil is subjected to sea-bottom pressures, and too much leveling produces burrs on the edges of the laminations with a consequent increase in the a-c resistance of the coil.

The core stacks are assembled on the winding in pairs with 0.0007-inch paper gaps and held together with phosphor bronze straps under a tension of 75 pounds (lower right in Fig. 7). A space of about 1/32 inch is maintained between the stack pairs to provide a cushioning layer of impregnating compound.

The core and winding are impregnated as a unit and then assembled in the metal tube with the liners holding them in place. After soldering on the winding terminals, the end plugs with their cable stubs are put in the tube and the whole is given a final impregnation.

The impregnating compound is a thermosetting, polyester resin, which is a thin liquid when its two components are first mixed, but after curing for several hours at 250 F (degrees Fahrenheit) assumes a consistency of rubbery gelatine. The function of the compound is, in addition to holding together the laminations and the winding turns, to float the core and winding inside the steel case, and to transmit sea-bottom pressures uniformly to all parts of the core, so that no bending stresses are set up in the laminations with a consequent loss in the coil inductance and an increase in the a-c resistance.

For the desired uniform pressure distribution, there must be no voids in the impregnated unit. Voids may occur as

Fig. 6 (above). Loading coil components

Fig. 7 (right). Loading coil construction details

a result of incomplete impregnation, shrinkage of 2.5% by volume during the curing process, and thermal shrinkage of about 4% between the 250 F curing temperature and normal room temperature. A further thermal shrinkage of 1% occurs as the coil goes to sea-bottom temperature of about 34 F.

Initial voids are minimized by careful vacuum impregnation after a thorough dehydration of the coil components and the compound itself. Curing and thermal shrinkages, except for the last 1% at sea bottom, are taken up by forcing the end plugs about 1/16 inch further into the coil case after impregnation. The set screws holding the end plugs in place are not finally seated until after this operation. Sixteen 1/4-inch holes in the coil case facilitate the impregnation, and later, under sea-bottom pressure, permit easy entry of enough of the jacket material to fill any small remaining voids without damage to the jacket.

After completing the impregnation, two 45-foot lengths of standard, insulated cable core are electrically brazed to the short stubs in the end plugs. The jacket is built up with a tightly wrapped butyl-synthetic rubber tape which forms into a compact, elastic mass without any heat treatment. The jacket is 5/16 inch thick over the coil, and extends several inches beyond the ends of the coil to make a water-tight joint with the insulation on the long cable stub. At this point the coil is given a 24-hour pressure test in water at 7,800 psi (pounds per square inch) to check insulation and effect of pressure on the inductance and a-c resistance.

Before armoring, the stub on each side of the coil is built up with burlap tape to provide a smooth taper extending to 20 feet on each side of the coil. This is followed by a wrapping of jute over the entire length to provide a bedding for the armor. The armor consists of 24 galvanized steel wires, 0.095 inch in diameter, extending the full length over the coil and cable stubs. Over the coil are interposed 39 additional wires, 0.112 inch in diameter, which are terminated, 13 at a time, at approximately 4 12, and 20 feet on each side of the coil as required to cover the tapered section. Armor wire ends are held in place by whippings of 16-gauge steel wire. Finally, over all, is the standard serving of tarred hemp to provide some mechanical protection for the armor. The finished diameter over the loading coil is 2.85 inches, and over the cable stubs about 1 inch.

The first lot of loading coils was armored at Cambridge, Mass., on a machine which required that the added wires over the coil be laid on by hand. Later coils were armored on a new and completely automatic machine, developed at the Newington, N.H., plant of the Simplex Wire and Cable Company, to armor the repeaters installed in the recent transatlantic telephone cables.

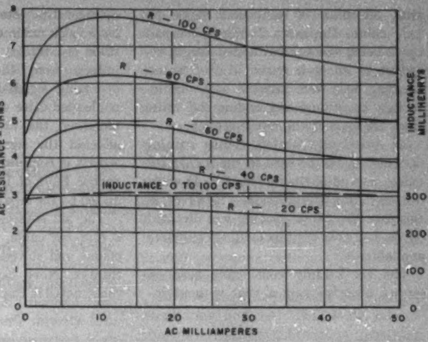


Fig. 8. A-c characteristics of loading coil

Coil Characteristics

The inductance of the loading coil is set at 300 millihenrys, at 60 cps, and 10 ma, by adjustment of the gap in one of the four sections of the core. There is normally little or no change in inductance between atmospheric pressure and sea-bottom pressure, provided that the leveling of the gap faces has been done to the proper degree, and that in the impregnation there have been left no appreciable voids in the vicinity of the core and winding. The a-c resistance is usually 20 to 25% greater at sea-bottom than at atmospheric pressure, but in most of the coils so far built the a-c resistance has been 15 to 25% less than the design value. As a result, the attenuation of the lump-loaded repair cable will be somewhat less than indicated by Figs. 1 to 4.

Fig. 8 shows the variation of inductance and a-c resistance with current for a typical loading coil under 8,000 psi pressure. The a-c resistance of the loading coil is normally 1 to 2 ohms less than the 7-ohm design value, and for currents up to 10 or 15 ma the increase in inductance and a-c resistance with current is slightly less in the loading coil than in the loaded cables. Above 15 ma the a-c resistance of the coil drops, while that of the cables increases rapidly with current. The inductance of the cables also increases rapidly at the higher current, but the inductance of the loading coil is nearly constant. While the inductance and a-c resistance, particularly at the higher signal currents, do not have the same pattern of change in the loading coil and in the cables, the increase in attenuation due to the higher currents is less in a coil-loaded repair section than in the loaded cable itself. In any case, the Hammel-Horta cable is the only one in which the signal currents exceed 15 ma in sections where the coils may be used. In this cable the currents are as high as 25 ma in the Hammel end, and 40 ma in the Horta end.

Under normal operating conditions, direct currents in the loaded cables due to earth potentials do not exceed 10 ma, and have very little effect on the inductance or a-c resistance of either the loading coils or the loaded cables. Direct currents up to 80 or 100 ma may flow in a cable circuit when protective devices installed at the terminals are operated by excessive earth potentials. Such currents will not damage the loading coils.

To check the effect on the loading coil of stresses encountered in laying or picking up the cable in deep water, one completely armored loading coil with long cable stubs was tested by being passed several times over a 7-foot-diameter drum while under longitudinal tensions up to 16,000 pounds. There was no mechanical damange to the coil, and the electrical characteristics were not affected.

Loading Coils in Use

In April 1956, the Hammel–Bay Roberts loaded cable (*2 HM-BR*) was interrupted 325 miles from Bay Roberts at a depth of 2,000 fathoms. Because of the poor condition of the cable (probably from damage sustained in the 1929 earthquake), 23 miles of cable had to be replaced. The replacement was made with 30 miles of nonloaded stock repair cable, the extra 7 miles of repair cable being the normal added length required to effect a repair in deep water. Loading coils were inserted at 5-mile intervals beginning and ending 2.5 miles from the ends of the repair section.

As previously stated, 5-mile spacing of the coils can be used in the *2 HM-BR* cable, instead of 3-mile (see Fig. 1) without degrading transmission. However, in this case the 5-mile spacing was dictated by the fact that only six loading coils were available.

The six loading coils have been in service now for over a year without incident. No attenuation measurements have been made on the *2 HM-BR* cable since the repair, but, as far as can be judged from the appearance of the operating signals, the insertion of the nonloaded cable and loading coils has not increased the attenuation or otherwise affected the signals. A comparison of signals is complicated by the fact that a partial fault of long standing was eliminated by the repair.

The six loading coils installed in the *2 HM-BR* cable, plus six recently completed and stored at the cable depot in Halifax, Nova Scotia, Canada, were all constructed in the Western Union Laboratory, New York, N.Y., except for the jacketing and armoring, which was done by the Simplex Wire and Cable Company. Arrangements are being made with a commercial supplier to manufacture a number of loading coils that will provide adequate stock for future loaded cable repairs.

The loading coils, complete with armor, cost about $2,000 each. Thus, for the repair of continuously loaded cables, the use of nonloaded stock repair cable with inserted loading coils, instead of continuously loaded repair cable, results in a saving of more than $1,000 per mile of repair cable used.

The Development of Television in Canada

J. E. HAYES
NONMEMBER AIEE

TELEVISION BROADCASTING began in Canada in September 1952, when the first two television stations owned by the Canadian Broadcasting Corporation (CBC) went into service in Montreal and Toronto. These stations were built with a loan of $4,500,000 provided by the Government of Canada for the purpose of establishing the nucleus of a Canadian television system. In the 5 years which have passed from that time to the present, the television industry has grown at a rate which far exceeded all expectations, and already about 80% of the population of Canada has television service available. The number of transmitting stations has increased from the original 2 to 39, of which 8 are owned by the Canadian Broadcasting Corporation, and 31 by private interests. In addition, 8 more private stations are under construction and several other applications are pending. CBC now has television stations in Halifax, Montreal, Ottawa, Winnipeg, and Vancouver carrying English network service, and second stations in Montreal and Ottawa carrying French network service.

One factor which promoted the rapid spreading of television stations to all parts of the country was the Government single-service policy which directed that new television stations should be built only in locations not already receiving a Canadian television service. This resulted in the construction of privately-owned television stations in most major cities from coast to coast, and prevented the grouping of a number of stations in the most densely populated areas where the commerical returns might be expected to be highest.

Television Networks

CBC operates two television network services, one in French and one in English, and all CBC and privately-owned stations carry one or the other of these services for part of their program schedules. CBC has entered into contracts with the Bell Telephone Company of Canada, the Canadian National Railway Company, and the Canadian Pacific Railway Company for the provision of microwave service from coast to coast. About 2,000 miles of microwave networks are now in operation, and when the service is completed, in 1958, to St. John's and Newfoundland in the East, and Victoria and British Columbia in the West, it will be well over 4,000 miles in length.

Film Recording

The original two CBC stations in Montreal and Toronto were equipped for making television film recordings, and as the number of CBC and privately-owned stations increased, the recording facilities were expanded so that they were capable of supplying sufficient film recordings to feed these television stations until microwave network service would permit direct connection. At the present time, approximately half of these stations are on the network, and the other half are being fed by film recordings.

The single-system direct-positive method of making film recordings was used as long as only one or two copies were required. When the need for several copies developed, the system was modified to photograph the picture as a negative so that any desired number of prints could be made. The sound is recorded on synchronous magnetic film and then directly rerecorded optically on each print. An electronic step-wedge generator is used to set up a 4-step gray scale on the kinescope of the film recorder, and a modified Densichron photometer checks the brightness of the steps. Densitometer measurements are then made of this gray scale on the negative and prints, thus facilitating close control of each step of the overall photographic process. In addition to this, *1B* sensitometric strips are attached to negative and positive film going through the processors and subsequent densitometer readings taken to assist in precise control of development. At present, CBC is using film at the rate of about 40 million feet per year, most of which is for film recordings. It has been found that the control procedures outlined above are necessary to ensure film recordings of consistently good quality.

Production Facilities

While the facilities for the distribution of programs were being expanded, the need developed for increasing the number of hours of service, and it became necessary then for CBC to build additional studio facilities, particularly in Montreal and Toronto. Montreal is the center of French network production, and Toronto the center of English network production, but CBC stations in other cities make contributions to the English network service on a regular basis. The English language network service now provides about 53 hours per week, of which 63% is Canadian production, and the remainder is imported partly from the British Broadcasting Corporation but mostly from the United States.

Studio facilities which are required to handle these programs are out of proportion to the percentage of Canadian content, because most of the television programs imported from the United States must have Canadian commercial announcements inserted in them, and many of these commercials are produced live in the CBC studios.

The French network, which at present consists of two CBC stations and three privately-owned stations, is fed about 57 hours per week, 69% of which is Canadian production. All CBC stations produce some programs for local service and so increase their total service to about 70 hours per week.

The present hours of service have been achieved by a continual growth in the studio facilities in Montreal and Toronto. In each city the CBC began with two studios, one mobile unit, and one film chain. These have now been increased to five studios, plus one under construction, two mobile units, and five film chains. In addition, an extra studio is planned for Montreal to permit increasing the French network service to the same number of hours as the English service. Even with these facilities much crowding is necessary to produce the number of programs on the present schedule.

The increase in number of studios has, of course, been accompanied by a large increase in the number of staff, and the area required for auxiliary services. This may be illustrated by the Toronto installation, where television broadcasting started out in a completely self-contained five-story building, which provided a total floor area of 52,000 square feet. This building housed the two studios, transmitter, film recording and reproducing facilities, administration and production offices, scenery shops, and storage space. The CBC has since rented additional space of about 70,000 square feet for storage, scenery shops, and dry rehearsal space, and 100,000 square feet for office space. The Toronto television staff which began with about 150 persons, now totals approximately 1,100.

System Design

The very considerable expansion in the amount of technical facilities has resulted in some very definite changes in system design, and which are being carried into the planning for the more recent studios. The original arrangement was based on feeds from the two studios, microwave from the mobile unit, film reproduction, and remote inputs feeding directly into master control and being switched from there to the transmitter, the recording room, and to the network once it became available. Initially, the film reproducing equipment and its control and monitoring facilities were contained in the master control room. While this arrangement worked satisfactorily for a short time, the activity in master control soon became so great that good operation became very difficult. It was decided to separate some of the functions from master control and to handle them in separate locations.

In the arrangement now in use, the basic idea followed is the divorce of all program production functions from master control, making it solely a distribution and quality control centre. To this end, a transmitter booth has been set up for each local station, and a utility booth used for production of film shows, remotes with local cut ins, and network shows with local cut ins which are intended for transmission onward to the network. Separate film reproducing rooms are used for originating film, and the film recording room remotely controls the selection of programs which are to be recorded. An examination of the work performed in each of these locations will help clarify the present mode of operation.

In the transmitter booth, program feeds are received via master control from the network, from local studios, from film chains, and from remotes. Programs are selected, commercials are inserted as required, and local spots and station identification are introduced, and all are fed via master control to the local transmitter.

In the utility booth, feeds similar to those provided the transmitter booth are available, and the same type of operation is performed with the output fed to the master control switcher for switching to the appropriate outgoing network. It is, therefore, possible for the local transmitter to be fed a different program from that which is being fed to a network. All slide and film feeds are supplied from the film reproducing rooms direct to master control, which routes them to a studio control room, a utility booth or a transmitter booth as required.

The recording room has remote controls for a special switching system located in the master control which permits the selection of up to eight programs which have been prepatched by master control. It is, therefore, possible for the recording room to record a sequence of several programs without reference to master control.

In addition to equipment for routing, monitoring, and controlling feeds, master control contains synchronizing generator and signal-locking facilities for the main plant. By means of a pulse switching system any desired combination of studios, film chains, and booths may be grouped on any synchronous generator, which can be synchronous locked to any incoming feed and so permit superimpositions and fades with remote and local signals.

Toronto master control has network feeds to and from the United States, eastern Canada, western Canada, and a separate feed to southwestern Ontario. Montreal master control has French and English network feeds to east and west.

In the production of elaborate variety shows, it has been found necessary to couple two and even three studios together for one program. For example, in Toronto a variety show may be produced in Studio 4, and the commercial produced in Studio 5. Both of these studios are remote from the main plant and each other, and a film insert which may form part of the commercial is fed to Studio 5 from the main plant.

The newest CBC studios are equipped with considerably more versatile vision mixers than were provided in the first studios. Complex programs now being produced require a mixer with eight input and four or five output circuits. Inputs frequently required are three cameras, one film chain, one remote studio or mobile unit, one network and one special effects feed. Outputs used are the main output, an effects output, a fade or dissolve output, and one or two preview outputs.

The complexities of the systems indicated above necessitate very close control of the performance characteristics of individual pieces of equipment involved, and in order to achieve this a considerable amount of test equipment is being provided and comprehensive maintenance and testing routines are being established.

Paper 57-665, recommended by the AIEE Television and Aural Broadcasting Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 28, 1957; made available for printing May 16, 1957.

J. E. Hayes is with the Canadian Broadcasting Corporation, Montreal, Que., Canada.

The author wishes to acknowledge the assistance of J. Carlisle, A. Ste.-Marie, and N. R. Grover in the preparation of this paper.

Audio Facilities

The need for more elaborate audio facilities has also become apparent. At the outset of television in Canada, CBC television studios were equipped with 24 microphone outlets. These were wired to a jack field which formed an integral part of the audio console. This console was equipped with 10 mixing potentiometers arranged so that groups of 3, 3, and 4 inputs were controlled by means of 3 submaster gain controls. The 10 inputs were also wired to the jack field and microphone selection obtained by patching at low level. As the program production pattern evolved, it became necessary to review the audio facilities. As an example, a typical program originating in a studio 60 by 90 feet calls for 17 microphones, 3 turntables, and 1 input from another studio from which the commercial portion of the program is produced. To cater to such requirements, studios now under construction will be equipped with more microphone outlets and an audio console with 15 mixing potentiometers with back-to-back switching allowing for 30 inputs to be set up without having to resort to patching. The outputs of the mixing potentiometers will go to a selector switching arrangement feeding 3 submaster gain controls. A low-level jack field will also be incorporated to give additional flexibility.

Color Television

The growth of color television in the United States, and the interest that has developed in Canada with respect to it, indicate that it may not be long before CBC may become involved in its complexities. With this in mind, considerable thought has been given to the possibility of conversion of some of the existing facilities. Studios at present under construction have been designed so that conversion may be facilitated. The number of technical problems which have arisen in the complex mode of operation undertaken to date in monochrome television have accentuated the need for very careful planning of color television facilities, if flexibility is to be provided and serious technical problems avoided. Planning to date shows the desirability of large developments, preferably at out-of-town sites, where sufficient area would be available to locate studios and their associated facilities in the most favorable arrangement. Color television requires much closer control of the transmission characteristics of all elements of the system, and imposes much more severe restrictions on the design of facilities than does the present monochrome television system.

Conclusion

The discussion so far has centered around the growth of the transmitting end of the television system. The picture would not be complete without some reference to the public's acceptance of television. The sale of television receivers has been even more rapid than the industry had anticipated. Already over 65% of Canadian families have television sets, and in areas where the service has been available for some time, the proportion is as high as 80%. Television service is now available to more than 80% of the population but as most of the larger cities and towns have now been covered extension of the service to the less densely populated areas will be a slower development. Some of this no doubt will be carried out by privately owned stations, but there will be many areas left where the economics of the situation will be such that private enterprise will not be interested in providing a service. It is quite possible that eventually CBC will be expected to supply television service to most of these areas as a part of its responsbility in providing a national broadcasting service.

Design Relationships for Iron-Core Filter Chokes

THERON USHER, JR.
ASSOCIATE MEMBER AIEE

WHEN FILTER CHOKES are designed for use in d-c power supplies, it frequently becomes necessary to make an estimate of the physical size of the choke, once the electrical ratings have been determined. The electrical ratings usually considered are the inductance, the d-c resistance, and the maximum allowable d-c current. Several fundamental relationships between these ratings and choke size are made. The relationships that result from the following analysis not only can be used for estimating the space requirements of a choke from a knowledge of the electrical ratings, but also point out interesting interrelationships of the electrical ratings.

Analysis

The following assumptions are inherent in the mathematical analysis that will follow.

1. The choke core contains standard *E-I* laminations.
2. The relative dimensions of the core are constant.
3. Shell-type construction is used.
4. The window area is completely filled with the winding. The space factor is constant, regardless of choke size.
5. The lamination stacking factor is neglected.
6. The effective permeability of the magnetic material and series air gap is constant.
7. The rated temperature rise of the choke is constant.

Assumptions 1 through 4, which concern the physical construction of the choke, are explained more thoroughly in Figs. 1 and 2, which show the dimensions of the core and winding. The standard dimensions of the *E-I* laminations are shown in Fig. 1. The core depth has been assumed to be twice the thickness of the outer leg of the lamination. This seems to agree reasonably well with the relative core depth for commercial chokes. From Fig. 2, the winding is seen to be perfectly symmetrical with circular corners. The lamination height, *h*, as shown in Fig. 1, will be used as the measure of the

---

Paper 57-743, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted July 24, 1956; made available for printing May 17, 1957.

THERON USHER, JR., is with Yale University, New Haven, Conn.

physical size of the choke in the final results.

Nomenclature

b = outer-leg width of E-I lamination, inches
h = height of E-I lamination, inches
L = choke inductance, henrys
μ = effective permeability
μ_i = incremental permeability of iron
α = air gap, inches
l_i = mean length of flux path in core, inches
A_i = cross-sectional area of core, square inches
N = winding turns
R = d-c winding resistance, ohms
ρ = resistivity of copper, (ohm-in) ohmn-inches
l_w = total winding length, inches
a_w = area of wire, square inches
l_t = mean-turn length, inches
k_{sp} = winding-space factor
A_w = total window area, square inches
I_r = rated current for rated temperature rise, amperes

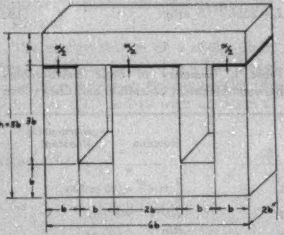


Fig. 1. Relative core dimensions

J_r = rated current density, amperes per square inch
P_c = copper loss, watts
W_s = surface loss per unit area, watts per square inch
P_s = total surface loss, watts
S = effective surface area, square inches
T = temperature rise above ambient, degrees centigrade
U_r = energy storage of choke for rated conditions, joules
VA = volt-ampere rating, volt-ampere (va)
τ = time constant of choke, inductance to resistance, L/R seconds (sec)

INDUCTANCE

The inductance of an iron-cored inductor with an air-gap is given by equation 1,[1] where L is the inductance in henries, μ is the effective permeability of the air gap and core in series, A_i is the cross-sectional area of the core in square inches, l_i is the mean length of the iron path in inches, and N is the number of turns.

$$L=3.2(10^{-8}\mu N^2 A_i/l_i) \qquad (1)$$

The effective permeability of the iron and air gap in series is given by equation 2,[2]

$$\mu=\frac{1}{1/\mu_i+\alpha/l_i} \qquad (2)$$

where μ_i is the incremental permeability of the iron, and α is the air-gap distance in inches. For assumption 6 to hold, l_i/α must be much smaller than μ_i, and this ratio must be constant for all inductors. Equation 2 then reduces to

$$\mu=l_i/\alpha \qquad (2A)$$

Since assumptions have been made concerning core dimensions and construction, equation 1 may be simplified by substituting expressions for the cross-sectional area and the length of the iron path into equation 1. The cross-sectional area is

$$A_i=(2b)(2b)=4b^2 \qquad (3)$$

The mean length of the iron path is

$$l_i=12b \qquad (4)$$

If the effective μ is given the constant value 100, equation 1 becomes

$$L=1.07(10^{-6}N^2b) \qquad (5A)$$

or

$$L=2.1(10^{-7}N^2h) \text{ henrys} \qquad (5B)$$

The assumed value for μ is much lower than μ_i for many core materials. For example, the incremental permeability of silicon steel varies roughly from 500 to 5,000 over a wide range of flux densities. Thus the inductance of the choke is controlled by the air gap, and the flux density in the iron need only be below the saturation value.

Equation 5(B) will be combined with later relationships to eliminate the factor involving the number of turns.

D-C WINDING RESISTANCE

The basic relationship for the resistance of the winding is found in equation 6.

$$R=\rho l_w/a_w \qquad \rho=6.8(10^{-7}) \text{ ohm-inches for copper} \qquad (6)$$

The symbol a_w is the wire cross-sectional area in square inches, and l_w is the total length of the winding in inches. The total length of the winding is related to the mean turn length in equation 7, and equation 8 gives the mean turn length in terms of the winding dimensions from Fig. 2.

$$l_w=Nl_t \qquad (7)$$

$$l_t=8b+\pi b=11.14b=2.23h \qquad (8)$$

Equation 9 expresses the wire area as a function of the window area and the space factor:

$$a_w=k_{sp}A_w/N \qquad (9)$$

From Fig. 1

$$A_w=3b^2=\frac{3}{25}h^2 \qquad (10)$$

If $k_{sp}=0.5$, the combination of equations 7 through 10 with equation 6 results in the final expressions for d-c winding resistance:

$$R=5.05(10^{-6}N^2b^{-1}) \qquad (11A)$$

or

$$R=2.5(10^{-6}N^2h^{-1}) \text{ ohms} \qquad (11B)$$

The assumption of a space factor of 0.5 is fairly reasonable, since bare wire stacked squarely yields a space factor of 0.785 and extra space must be allotted for insulation and layer separators.

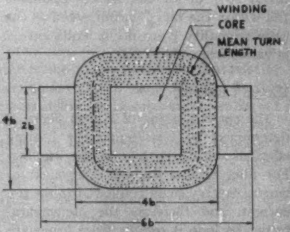


Fig. 2. Cross section of choke

RESISTANCE TO INDUCTANCE RATIO

Equations 5(A) and 11(B) give the inductance and resistance as functions of the number of turns and the lamination height. It is desirable for size estimation purposes to obtain an expression which does not depend upon N. This is easily done by dividing equation 11(B) by equation 5(B). The resistance to inductance ratio thus determined depends only upon the lamination height and is found in equation 12.

$$R/L=120/h^2 \qquad (12)$$

If the resistance and inductance of a choke are known, equation 12 will give the approximate size. Moreover, if the size of a choke is specified, the transient response and frequency response are specified to some degree, since L/R is the time constant of the choke.

In Fig. 3 the size, h, is plotted as a function of R/L for actual filter chokes manufactured by several concerns.

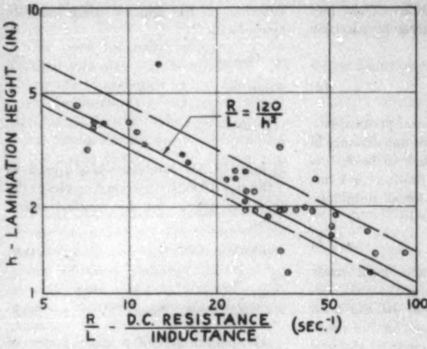


Fig. 3. Lamination height versus R/L ratio

Equation 12 is also shown on Fig. 3. For actual chokes, there is some deviation from the idealized characteristic, but about 80% of all the chokes examined fall within the range encompassed as the proportionality constant in equation 12 is varied from 100 to 200.

RATED CURRENT

The current rating of a filter choke is determined by considering the heating effect of the d-c current in the winding. For filter chokes the a-c component of current is very small compared to the d-c current, and therefore the incremental hysteresis loop is small. Core losses are thus negligible compared to d-c winding losses, and the final temperature rise is determined by the latter.

The current in the winding is related to the current density by equation 13.

$$I = Ja_w \tag{13}$$

When equations 9 and 10 are combined with equation 13, equation 14 is obtained as follows

$$I = 1.5Jb^2/N = 0.06Jh^2/N \tag{14}$$

The power dissipated in the windings is found from equations 14 and 11(B):

$$P_c = RI^2 = 11.3(10^{-7}J^2b^3) = 9.0(10^{-8}J^2h^3) \text{ watts} \tag{15}$$

When the temperature of the choke reaches the equilibrium value, the winding loss equals the power dissipated at the surface of the choke. The surface loss per unit area is given by equation 16,[3]

$$W_s = (T/100)^{1.21}, \text{ watts per square inch} \tag{16}$$

where T is the temperature rise in degrees centigrade. The total surface loss for the choke is found by multiplying equation 16 by the effective surface area, S as given in equation 17

$$P_s = (T/100)^{1.21}S \text{ watts} \tag{17}$$

The effective surface area must be approximated from information given in Figs. 1 and 2. The area of all exposed surfaces is $113b^2$. The area of a sphere with a diameter equal to the mean outer choke dimension is $78b^2$. Also, the core of the choke will probably not be as hot as the windings. An effective area of $100b^2$ will be assumed as a compromise.

The rated current density is found in terms of choke dimensions by equating P_c to P_s for rated temperature rise:

$$9.0(10^{-8}J_r^2h^3) = (T_r/100)^{1.21}4h^2$$

$$J_r = \sqrt{\left(\frac{T_r}{100}\right)^{1.21}\frac{4.44(10^7)}{h}} \tag{18}$$

For $T_r = 45$ C, equation 18 becomes the following

$$J_r = 4{,}100h^{-1/2} \text{ amp/in}^2 \tag{18A}$$

The rated current is found by combining equation 14 with equation 18(A), giving equation 19:

$$I_r = 250h^{3/2}/N \text{ amp} \tag{19}$$

Table I. Summary of Basic Relationships Between Electrical Quantities and Choke Size

	Relation	Proportionality Constant
1.	$\frac{R}{L}h^2 = \frac{h^2}{\tau}$	$= 100$ to 200
2.	$J_rh^{1/2}$	$= 4{,}100$
3.	$\frac{RI_r}{\sqrt{L}}$	$= 4$ to 14
4.	U_rh^{-4}	$= 1.0(10^{-3})$ to $3.0(10^{-3})$
5.	$\frac{VA}{U_r^{0.5}}$	$= 250$

BASIC CONSTANT INVOLVING ELECTRICAL RATINGS

Filter choke specifications almost always contain the inductance, d-c resistance, and rated current of the choke. Equations 5(B), 11(B), and 19, which give these three quantities in terms of N and h, may be combined in a fashion so as to eliminate both of the latter variables from the result.

The quantity $RI_r/\sqrt{L}$ turns out to be a constant:

$$\frac{RI_r}{\sqrt{L}} = \frac{[2.5(10^{-6}N^2h^{-1})](250h^{3/2}N^{-1})}{4.6(10^{-4}Nh^{1/2})} = 13.5 \tag{20}$$

Calculations of $RI_r/\sqrt{L}$ for commercial filter chokes of widely varied ratings yield values which range from 4 to 14. The variation in the actual ratios occurs because the simplifying assumptions of

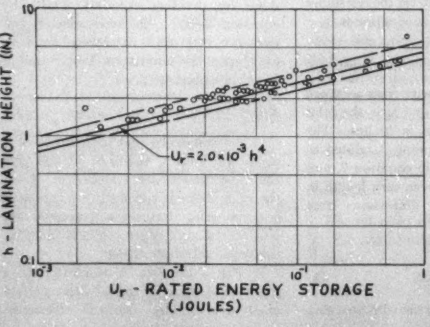


Fig. 4. Lamination height versus rated energy storage

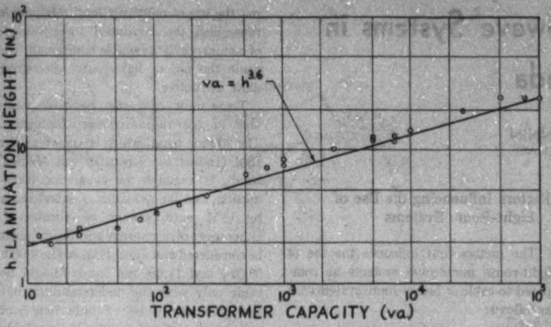


Fig. 5. **Lamination height versus volt-ampere capacity for single-phase 60-cycle transformers**

the theoretical analysis do not strictly hold from choke to choke.

Energy Storage and Equivalent *VA* Rating

The energy storage of a choke under rated conditions is found as a function of lamination height by combining the two equations 5(B) and 19 with equation 21 below.

$$U_r = \frac{1}{2} L I_r^2 \tag{21}$$

$$= \frac{1}{2}[2.1(10^{-7}N^2h)](250h^{3/2}N^{-1})^2$$

$$U_r = 6.4(10^{-3}h^4), \text{ joules} \tag{21A}$$

Fig. 4 portrays h as a function of U_r for a number of commercial filter chokes. The actual relationship is very close to a fourth power law, although the constant of proportionality is lower. Almost all filter chokes are included in the region where the constant of proportionality in equation 21(A) is allowed to vary from $1(10^{-3})$ to $3.0(10^{-3})$.

Quite often for engineering purposes, it is convenient to give an "equivalent volt-ampere rating" for a choke. This means that the physical size of a given choke is identical to that of a transformer with the specified *VA* rating.

The *VA* ratings for commercial single-phase 60-cycle transformers are plotted against lamination height, h, in Fig. 5. The data follows the empirical power law relationship in equation 22 fairly closely.

$$VA = 1{,}000(h/6.8)^{3.6} = h^{3.6} \tag{22}$$

Equation 22 may be combined with equation 21(A) to give the equivalent *VA* rating of a filter choke in terms of the rated stored energy. A proportionality constant of $2(10^{-3})$ is assumed for equation 21(A), so that closer approximation to the actual case is assured.

$$VA = 250U_r^{0.9} \tag{23}$$

Note in equation 23 that the relationship is almost linear.

Summary of Results and Conclusions

A summary of the basic relationships between important electrical quantities and choke size is given in Table I. The actual variational limits of the proportionality "constants" are also listed.

From the five basic relations given in Table I, practically any important variable concerning filter choke performance may be estimated quickly. Relation 1 combines the time constant and the size, and relation 4, the energy storage and the size. Relation 3 deals solely with the electrical ratings of the choke. Relation 2 is helpful in designing chokes, and relation 5 is helpful to the engineer who likes to think in terms of equivalent-sized single-phase 60-cycle transformers.

It is noted from the table that there is quite a bit of variation in the supposed "proportionality constants." This variation is due largely to the fact that relative lamination depth, space factor, effective permeability, and temperature rise are not strictly constant. However, the relationships have been derived in such a way as to allow estimates to be made with an error which is comparable to that of most preliminary engineering design estimates.

References

1. Radiotron Designers Handbook (book), F. Langford Smith, editor. Radio Corporation of America, Princeton, N. J., fourth edition, 1952, p. 242.
2. *Ibid.*, p. 245.
3. *Ibid.*, p. 238.

Light-Route Microwave Systems in Canada

A. J. DINNIN
NONMEMBER AIEE

Synopsis: This paper reviews the use of light-route microwave systems as auxiliaries to the TransCanada *TD*-2 microwave system for television and telephone transmission. It briefly outlines the reasons for their use, and discusses the need for diversity reception to obtain reliability. Specific applications along the *TD*-2 route are described and a review is made of possible future developments.

A LIGHT-ROUTE microwave system may be defined as one capable of carrying approximately 100 telephone circuits for distances up to 200 miles or more, or a single television video signal for somewhat shorter distances. Such systems may be used for independent routes or for end-links and spur routes on a long-haul system. Their useful field may be compared to the use of interurban and transcontinental trains in the railway business in that they fill the need for more economical and more easily operated systems where the service demands are not as exacting.

Light-route systems currently in use or planned along the TransCanada *TD-2* route are assigned frequencies in the 900-, 2,000-, 4,000- and 6,000-mc (megacycle) bands. The 11,000-mc band will probably be used to good advantage as equipment becomes available and experience is obtained on its operation.

Factors Influencing the Use of Light-Route Systems

The factors that influence the use of light-route microwave systems as compared to cable or heavy-route systems are as follows:

1. Relative Cost: Television circuits over about 5 miles in length may be provided more economically by microwave than by cable. This is especially true where underground cable construction is difficult and expensive. For telephone transmission, where a small number of circuits are required, light-route systems may be more economical than the heavy-route type.
2. Use of Existing Facilities: Where the route parallels that of an existing system, towers, buildings, power supplies, and auxiliary services may be shared.
3. Ultimate Requirements: Where present requirements are small but heavy future growth is anticipated, a light-route system may be employed temporarily to defer the expense of a heavy-route system.
4. Maintenance: In general, a microwave system, used for television transmission, will require less maintenance effort than a cable facility.
5. Transmission Characteristics: Better transmission quality may be obtained over a microwave system than over a long cable facility.

In addition to these factors, the recent development of ferrite isolators which permit the use of relatively long wave-guide runs, and the continued improvement of commercially available equipment, has made the use of light-route microwave quite attractive.

There must, of course, be space available in the radio-frequency spectrum. The 900-mc band, which is known as the ISM (Industrial, Scientific, and Medical band), is available for telephone transmission, and the question of interference by ISM must always be considered. Other users of microwave equipment must be considered and allocations in the 4,000-6,000-, and 11,000-mc bands should be made only with the understanding that they give way to heavy-route long-haul systems since these can use the available band width very much more efficiently.

Finally, consideration must be given to availability, compatibility, reliability, and ease of maintenance. The question of transmission reliability is discussed in the following paragraphs.

Diversity Reception

Reliability is a primary objective in the design and operation of a common-carrier transmission facility. Interruptions, even though momentary, are very undesirable in voice or program transmission, and should be eliminated or kept to the very minimum in the provision of circuits for teletypewriter, data, and other types of communication.

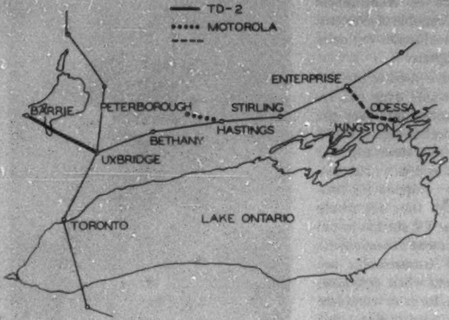
Paper 57-707, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 20, 1957; made available for printing May 17, 1957.

A. J. Dinnin is with The Bell Telephone Company of Canada, Montreal, Que., Canada.

Table I. Summary of the Light-Route Microwave Systems Associated with the TransCanada TD-2 Microwave System

System	No. of Sections	Total Length, Miles	Type of Equipment	Frequency Band, Mc	Use	TV Station Served or Type of Service Provided
Sydney Mines–Sydney	1	7.8	TD-2	3,700–4,200	TV	CJCB-TV Sydney, N.S.
Hardwood Hill–Churchill	1	44.9	TD-2	3,700–4,200	TV	CFCY-TV Charlottetown, P.E.I.
Dawson–Moucton	1	10.1	TD-2	3,700–4,200	TV	CKCW-TV Moncton, N.B.
Enterprise–Kingston	2	29.6	Motorola	5,925–6,425	TV	CKWS-TV Kingston, Ont.
Hastings–Peterborough	1	15.0	Motorola	5,925–6,425	TV	CHEX-TV Peterborough, Ont.
Uxbridge–Barrie	1	37.0	TD-2	3,700–4,200	TV	CKVR-TV Barrie, Ont.
North Bay–Callander	1	9.3	TD-2	3,700–4,200	TV	CKGN-TV North Bay, Ont.
Nellie Lake–Timmins	1	31.9	TD-2	3,700–4,200	TV	CFCL-TV Timmins, Ont.
Port William–Port Arthur	1	4.2	Motorola	5,925–6,425	TV	CFPA-TV Port Arthur, Ont.
Fort William–Longlac	6	174.7	Motorola	5,925–6,425	Telephone	Lenkurt 33B and 45BX Carrier
Regina	1	5.0	Motorola	5,925–6,425	TV	CKCK-TV Regina, Sask.
Bowell–Medicine Hat	1	11.4	Motorola	5,925–6,425	TV	CHAT-TV Medicine Hat, Alta.
Edmonton	1	6.1	Motorola	5,925–6,425	TV	CFRN-TV Edmonton, Alta.
Lethbridge	1	2.3	Motorola	5,925–6,425	TV	CJOC-TV Lethbridge, Alta.
Hedley–Granum	10	328.9	RCA-CTR-900	890–960	Telephone	Lenkurt 33B and 45BX Carrier
Salmo–Nelson	3	69	RCA-CTR-900	890–960	Telephone	Lenkurt 33B and 45BX Carrier
Hedley–Vancouver	5	156.4	Lenkurt 72C*	890–960	Telephone	Lenkurt 33B and 45BX Carrier
Vancouver–Victoria	2	62	General Electric	1,700–2,300	TV	CHEK-TV Victoria, B.C.

* Lenkurt 72C will be replaced by Lenkurt 74A equipment operating in the 5,925–6,425-mc band.



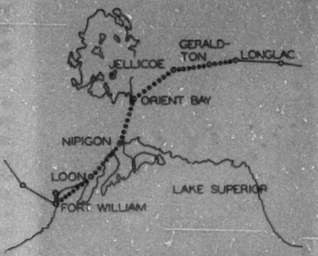


Fig. 1 (left). TV spur routes in southern Ontario

Fig. 2 (above). Fort William–Longlac Motorola system

Multipath fading is an important factor in the control of reliability. This type of fading is affected by transmitter power output, receiver sensitivity, antenna gain, and the characteristics of the path itself. In order to obtain the required reliability, within the limits of a reasonably economical design, it has been found necessary to protect against multipath fading by using some form of diversity reception. This is particularly necessary at the higher microwave frequencies and on overwater paths, and is more important for telephone than for television transmission.

There are two types of diversity reception. In frequency diversity reception two radio channels operate simultaneously, but on separate frequencies, over the same path. Both channels are modulated by the signal to be transmitted, and an automatic switch or receiver combining unit is used to switch from one receiver output to the other as dictated by the received signal level. Space diversity reception requires two receiving antennas separated vertically some 50 to 100 feet or horizontally about ten times as much, the exact spacing depending upon frequency and the repeater spacing.

Frequency diversity is to be preferred over space diversity for several reasons. In the first place the frequency-diversity arrangement results in automatic protection of all radio equipment while in space diversity the transmitter must be protected as a unit on a switch-over basis. Secondly, space diversity requires a duplicate receiving antenna system and a higher or duplicate tower for its support, thus making it more expensive than frequency diversity. While it is true that frequency diversity requires twice as much space in the available band of frequencies, as space diversity, no one is more conscious of the value of frequency space conservation than those who operate radio in the band. It would not be out of order to expect that the users of microwave radio would make an effort to coordinate paralleling systems and, if necessary, change from frequency to space diversity operation or make other rearrangements, in situations where more urgently needed requirements for frequencies arise.

Light-Route Microwave Applications

The locations where light-route microwave systems are now existing or are planned along the TransCanada *TD-2* route together with pertinent information and data, are summarized in Table I. These systems will be briefly discussed and for this purpose will be divided into two groups covering first those systems used for television transmission and second those used to provide telephone circuits.

TELEVISION

The *TD-2* systems feeding the television stations at Sydney, Moncton, Barrie, and Timmins are single-channel one-way systems operating without protection, and each consists of one repeater section. The antennas are parabolas 8 feet in diameter and are fed by means of rectangular wave-guide. In the first three of these systems the antennas are mounted on towers or buildings for direct transmission to the distant station, while on the Barrie and Timmins systems they

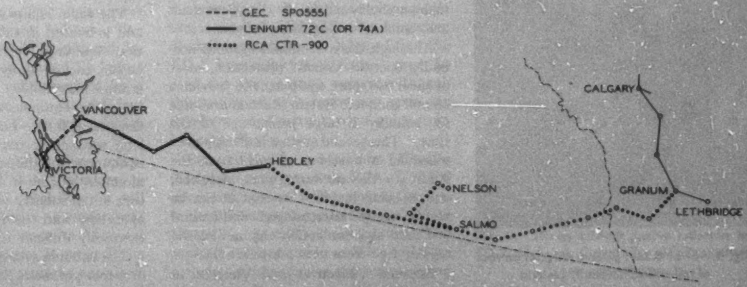


Fig. 3. Light-route systems in British Columbia and Alberta

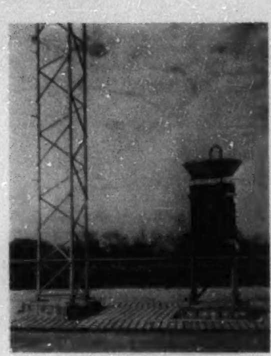


Fig. 4. Motorola microwave terminal on roof of toll office, Kingston, Ontario

are located on the ground and illuminate 10 by 15 foot passive reflectors mounted on suitable towers.

The Hardwood Hill–Churchill *TD-2* system consists of a single repeater section. It uses 8-foot parabola antennas mounted on towers and fed by means of rectangular wave-guide in a standard manner. The transmission path is rather long (45 miles) and is mostly over water, so that protection from multipath fading is essential. For this reason two radio channels separated by 160 mc were installed and equipped with an automatic switch to provide frequency diversity reception.

Each of the above *TD-2* systems have been designed with economy in mind and yet with no sacrifice of transmission quality. They are all capable of excellent transmission of monochrome or color television program signals.

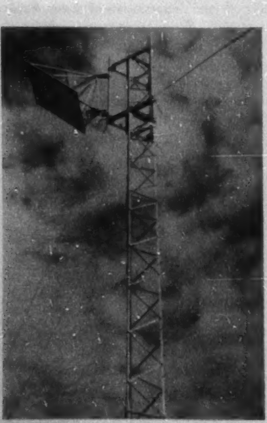


Fig. 5. 6- by 8-foot passive reflector on roof of toll office, Kingston, Ontario

Motorola microwave is used to provide a single television channel over two repeater sections to Kingston and over one repeater section to Peterborough. Fig. 1 shows the layout of these systems together with the Barrie *TD-2* system. The two Motorola systems are equipped for two-way operation although they are rarely used for feeds from the TV station to the network. The equipment components for this direction of transmission are therefore used as spares when available, and protection in the form of complete terminals is not provided. Since this equipment is now several years old, extensive modification would be required to permit satisfactory color television transmission.

This Motorola equipment is not equipped with ferrite isolators and long wave-guide runs are not permitted. For this reason remote outdoor radio-frequency cabinets or composite repeater stations are used so that antennas may be mounted as close as possible to the transmitting klystrons. The antennas are all 6-foot parabolas and they illuminate tower-mounted passive reflectors which are 8 by 12 foot in size except for one 6 by 8 foot in size at Kingston.

At Port Arthur, Regina, Medicine Hat, Edmonton, and Lethbridge, Motorola microwave systems are also being used for television transmission. They are all single repeater section systems operating over relatively short paths which are in the order of 5 to 10 miles long. The radio equipment is of recent design and has suitable transmission characteristics for the transmission of color as well as for monochrome television. Ferrite isolators are standard equipment in this design permitting the use of relatively long wave-guide runs and thus making it possible to mount the equipment on suitable racks indoors. In general the antennas are 40-inch parabolas and at a few locations these are combined with passive reflectors.

Although these Motorola systems normally operate without protection, most of them use spare equipment to provide a second complete system which is available for standby a large percentage of the time. This second system is usually provided for a television channel from the studio to the network, when required, and is being installed so that it can be rather easily interchanged end-for-end and used as a protection channel for the regular feed from network to studio.

Between Vancouver and Victoria, in British Columbia, the Northwest Telephone Company has installed a two-repeater section system which was manufactured in England by the General Electric Company. This system operates in the 1,700–2,300-mc band and uses coaxial cable to feed 12 foot parabolic antennas. The choice of this equipment was influenced by the relatively long over-water transmission paths of 38 and 24 miles respectively. Protection is not provided for this system.

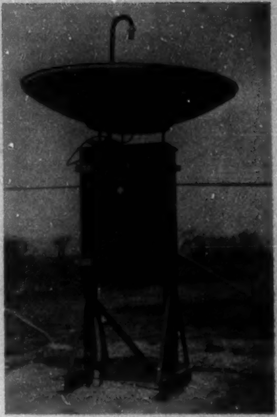


Fig. 6. Motorola microwave terminal equipped with 6-foot parabolic antenna at Hastings TD-2 Auxiliary Repeater Station

TELEPHONE

The Fort William–Longlac system consists of six repeater sections. Fig. 2 shows the route layout. It employs Motorola microwave equipment operating in the 5,925–6,425-mc common-carrier band with a capacity of 120 voice-frequency channels and it employs frequency diversity reception.

The radio equipment is rack-mounted and is housed in the *TD-2* repeater station buildings where commercial a-c power, backed up by diesel-powered generators, is available. Indoor rack-mounting of the equipment is possible through the use of ferrite isolators which permit long wave-guide runs between the radio equipment and the antenna. Equipment mounted indoors is a very desirable feature in the country where outdoor winter operation and maintenance would be extremely difficult.

The antenna system consists of 40-inch or 6-foot parabolas located at the base of

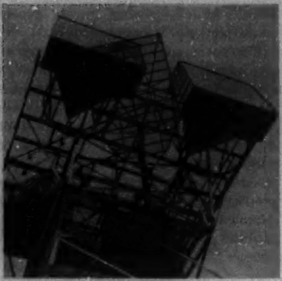


Fig. 7. 8- by 12-foot passive reflector on TD-2 tower at Hastings TD-2 auxiliary repeater station

the *TD-2* towers and illuminating 10 by 15 foot (8 by 12 foot at one location) passive curved reflectors mounted on these towers. Assuming 100 milliwatt transmitter power output, the calculated received signal strength for the six repeater sections lies between −40 and −50 decibels referred to one milliwatt. Although the transmission paths are rather long and one of them is an over-the-water path, this allows a reasonable margin for fading and deterioration of equipment. Ultimately it is expected that the present antenna system will be removed and arrangements made to use the *TD-2* horn-reflector antenna instead. This rearrangement may be dictated by a requirement to remove some of the *TD-2* tower loading or to improve the operation of the system and permit more economical use of the frequency band. It will be possible when suitable wave-guide combining networks become available.

Frequency diversity reception, as used on this microwave system, switches the protection channel into service in each repeater section independently of other sections. This is different from the *TD-2* automatic switching system which switches the protection channel over several repeater sections but it provides maximum protection against outages due to fading or equipment failures. The operation of the switch is initiated by interruption of a pilot, caused by failure of some equipment component or by a change in automatic gain control bias resulting from a fade in the transmission path.

The use of frequency diversity reception was considered to be necessary to provide the required reliability for toll circuit operation. Space diversity was studied in the design stage but was not accepted because, to use it properly on this route, would have required expensive tower rearrangements. Furthermore space diversity requires switch-over operation of transmitters and suitable arrangements for this were not available at the time this system was engineered.

Two types of multiplexing equipment are combined to provide the required telephone circuits. An *H* carrier system and the voice frequency band beneath it provide for the *C1* alarm and order wire circuits required to serve each *TD-2* repeater station. Above the *H*, a Lenkurt *45BX* carrier provides toll circuits to Fort William, Nipigon, Geraldton, and Longlac. A video decoupler unit permits dropping and reinserting the *H* carrier and voice channels at each repeater section without interrupting the Lenkurt *45BX* carrier system.

In British Columbia, the Northwest Telephone Company is planning to use Lenkurt *72C* which operates in the 900-mc band (or *74A* which operates in the 5,925-6,425-mc band) and RCA *CTR-900* (another 900-mc system) to provide the necessary *TD-2* auxiliary circuits through the Rocky Mountains and in addition a number of toll circuits. Fig. 3 shows the layout of these systems.

From Vancouver to Hedley, a distance of five *TD-2* repeater stations, the initial installation will be Lenkurt *72C*. This system operates in the 900-mc band and has a capacity of 120 telephone circuits on each of up to three radio channels. The equipment will be housed in the *TD-2* radio repeater stations and the 10-foot parabolic antennas, fed by means of coaxial cable, will be mounted on pole structures rather than on the *TD-2* towers. Pole mounting will permit replacement of the system at a later date without interruption to service. The antennas will be equipped with heaters to minimize the effect of icing which may be quite severe at some locations.

Consideration is being given to the replacement of the Lenkurt 900-mc system by a Lenkurt 6,000-mc system when it becomes available. The *74A* will have a capacity of 240 telephone circuits per radio channel and will use frequency diversity reception. It is hoped that the *TD-2* horn-reflector antennas may be used for the antenna system although this arrangement is contingent upon the development of suitable wave-guide combining networks.

Fig. 8. 40-inch parabolic antenna and wave-guide supporting structure at Orient Bay TD-2 auxiliary repeater station

Between Hedley and Granum (Alberta) it is planned to install an RCA *CTR-900* microwave system. This system operates in the 890-960 mc band and has a 4-150-kc baseband. It will consist of ten *TD-2* repeater sections and in addition will have a three section spur to the alarm center at Nelson. The installation will be similar to the *72C*, as in the foregoing, in that it is installed in the *TD-2* stations and uses coaxial cable to feed the 10-foot antennas mounted on the *TD-2* towers. All antennas will be equipped with heaters.

Two types of multiplex will be used on these radio systems. A Lenkurt *33B* multiplex will provide three circuits for the *C1* alarm and order wire which is required to appear at each repeater station, and also for the automatic switch control circuits which appear at the ends of switching sections. Above the *33B* a Lenkurt *45BX* system will provide for

Fig. 9. 10- by 15-foot passive reflectors on TD-2 tower at Orient Bay TD-2 auxiliary repeater station

the additional circuits for order wire and control and for toll service.

Figs. 4–9 show various installations of microwave equipment.

Future Developments

Although over-the-horizon or scatter systems might be considered as competitive to light-route systems, it is felt that they are only competitive where it is not possible to install a conventional microwave system. This might happen where the route lies over country where there are no roads and no commercial power, and where there is no requirement for circuits to locations along the route itself. One disadvantage of their use is the possibility of interference to ultrahigh-frequency television transmission.

The development of a light-route system operating in the 10,700-11,700-mc common carrier band is now in progress in the United States. This system should prove to be very useful in areas where interference with systems operating in the lower frequency bands is a problem. It will have a capacity of 240 telephone circuits for use on paths up to ten repeater sections in length.

The use of light-route systems is attractive where a relatively small number of circuits or a television channel is required. As the demand for circuits grows they will likely be replaced by heavy-route type of system in order to use the frequency spectrum in the most economical manner. Their use may be expected to increase in proportion to any decrease in first cost or any increase in the cost of land line facilities and the relative costs of maintenance.

Some Relations Between Television Picture Redundancy and Bandwidth Requirements

K. H. POWERS
NONMEMBER AIEE

H. STARAS
NONMEMBER AIEE

Synopsis: An analysis is presented of the concept of statistical redundancy. It is shown that the redundancy can be separated into two parts, nonpredictive redundancy and predictive redundancy. The nonpredictive redundancy results from the message not possessing the most efficient first-order probability distribution while the predictive redundancy results from the statistical correlation of the present value of a signal with its past history.

The analytical procedure presented in this paper is then applied to the television communication problem. The results of this analysis compare very favorably with existing experimental evidence. The analysis indicates that by clever statistical coding, a bandwidth reduction of about two to one might be obtained, although even this modest gain would require complex equipment. The conclusion is reached that the best way of achieving bandwidth reduction in television communication is by cleverly degrading the information in a picture in such a way that the human observer would not notice it appreciably.

THE large bandwidth required for the transmission of standard television signals has prompted a study of the rate at which information in a televised scene is displayed to the human observer. Because of the limited spatial resolution, brightness sensitivity, and recovery time of the eye, the observer cannot perceive with precision all of the information in a televised scene. To reproduce an image with more accuracy than necessary is to supply redundant information. This type of redundancy is virtually independent of the characteristics of the scene and may be classed as nonstatistical redundancy. The adoption of sequential framing, line scanning, and limited horizontal resolution in the present television standards has to a large degree minimized this redundancy.

Statistical Redundancy

A second type of redundancy in television is manifest by the statistical correlations that exist between the elements of a given picture or between corresponding elements of successive frames. It is this statistical redundancy with which this paper is concerned.

Consider the standard scanning television system in which the televised scene is converted into a continuous time-varying waveform, limited in frequency to a video bandwidth of 4 mc (megacycles). By the sampling theorem, such a waveform is completely specified by giving its values at samples points spaced at the Nyquist interval, 1/8 microsecond. The television signal may thus be represented by an infinite sequence of random variables. Because of the limited brightness sensitivity of the eye, it can further be assumed that each element of the sequence can assume only one of a finite set of r values. Subjective tests have indicated that somewhere between 32 and 100 halftone levels are sufficient for reasonable reproduction. These values correspond, approximately, to 30- and 40-db (decibel) signal-to-noise ratios.

If the successive elements take values independently at random, and if the probability of occurrence of the ith level is $p(i)$, then the information conveyed by an element assuming the ith level is $-\log p(i)$. The average information conveyed by each element in the sequence is the entropy of the distribution given in the following equation

$$H=-\sum_{i} p(i) \log p(i) \tag{1}$$

The average information has its maximum possible value when each level is equally probable with probability $1/r$. This maximum value, $\log r$, is referred to as the source capacity and is denoted by C. The nonnegative quantity $(C-H)$ is called, for reasons that are apparent later, the nonpredictive redundancy.

Now if successive element values occur, not independently, but with statistical influence extending over past elements, the average information per element is less than that given in equation 1. If s_j denotes a state representing a particular past history, i.e., a set of known past values in the sequence, the average information per element is given by

$$I=-\sum_{i,j} p(i,s_j) \log p_{s_j}(i) \tag{2}$$

where $p(i,s_j)$ is the joint probability that

Paper 57-722, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 29, 1957; made available for printing April 24, 1957.

K. H. Powers and H. Staras are with the Radio Corporation of America Laboratories, Princeton, N. J.

state s_j occurs and is followed by an element assuming the ith level; $P_{s_j}(i)$ is the conditional probability that, given the occurence of the state s_j, the next element will assume the ith level.

The information rate is smaller in a correlated sequence because a knowledge of the past history provides some a priori information about the next element before that element occurs. In fact, one can evaluate the average quantity of information about the present element provided by a knowledge of the state s_j. It is, simply,

$$\sum_{i,j} p(i,s_j) \log \frac{p_{s_j}(i)}{p(i)} = H - I \qquad (3)$$

This information is called the predictive redundancy because it gives a quantitative measure of how well an element in the sequence can be predicted from a knowledge of the past history.

The total statistical redundancy, i.e., the difference $(C-I)$ between the capacity of the source and the actual information rate, can be expressed as the sum of the nonpredictive and predictive redundancies.

$$C-I=(C-H)+(H-I) \qquad (4)$$

It should be noted that the nonpredictive redundancy is quite independent of the statistical correlations between the elements, but gives a measure only of the deviation from uniformity (equally probable levels) of the first-order probability distribution.

Continuous Distributions

Consider the case wherein the elements of the sequence can take values from the real continuum according to a set of probability density functions. In such a case, each element is capable of conveying an infinite amount of information since the probability of occurrence of any specific value is zero. However, as it is impossible to measure a quantity with unlimited precision, because of imperfect instruments and the presence of noise, the average information conveyed is, in general, finite. Let the signal be represented by a sequence $\{f_i\}$ of elements $f_k = m_k + n_k$, where $\{m_i\}$ is the message sequence and $\{n_i\}$ is an error sequence, or noise, which is assumed to be gaussian in distribution and statistically independent of the message. The average information I provided by each element f_k about the message sequence $\{m_i\}$ has been shown by Shannon to have its maximum possible value when the message elements are also multivariate gaussian in distribution.

The statistical structure of a gaussian sequence is completely described by the set of coefficients of linear correlation between the sample elements, or, equivalently, by the power spectrum of the original time function. If the message and the noise are assumed gaussian, with power-density spectra $P(f)$ and $N(f)$, respectively, the average information I (per element) is given by the following equation

$$I = \frac{1}{2W} \int_0^W \log \left[1 + \frac{P(f)}{N(f)} \right] df \qquad (5)$$

where W is the message bandwidth. I has its maximum value C when the message spectrum $P(f)$ is so chosen that $P(f)+N(f)$ is constant. It may be shown readily that the predictive redundancy of the message sequence, i.e., the information about the present element provided on the average by the elements of the past, is given by

$$R = \frac{1}{2} \log \frac{1}{W} \int_0^W P(f)df - \frac{1}{2W} \int_0^W \log P(f)df \qquad (6)$$

which follows to be the limit, as the noise vanishes, of the quantity $(C-I)$. In the gaussian case, the predictive redundancy is the total statistical redundancy. This follows from the fact that the entropy of a continuous random variable of fixed power P has its maximum possible value when the first-order distribution of the variable is gaussian with variance P. Thus the nonpredictive redundancy of a gaussian sequence is zero.

For a more general, nongaussian, message with spectrum $P(f)$, equation 5 gives an upper bound to the information rate. Thus equation 6 represents a lower bound to the predictive redundancy. In the general case, however, the computation of information, hence of statistical redundancy, is a very formidable problem. It will be of value to compute the statistical redundancy of a typical television image under the gaussian assumption, but refer to the values obtained as the linear redundancy, that redunancy due to the linear correlations existing in the sequence. The linear redundancy is the redundancy that can be removed by means of linear prediction.

Values for linear redundancy have been computed for typical correlations as measured by Kretzmer.[1] These values are compared with similar computations of statistical redundancy in a simple nongaussian case, as well as with measurements obtained by Schreiber[2] on typical television images.

Linear Redundancy in Television

Let the coefficients of linear correlation of the stationary sequence $\{f_i\}$, having zero mean, be defined by

$$\rho_m = \frac{\overline{f_k f_{k+m}}}{\overline{f_k^2}} \qquad (7)$$

where the bar denotes the average or expected value. If the nth order correlation determinant is defined by

$$P^{(n)} = \begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & & \\ \vdots & & & \vdots \\ \rho_{n-1} & \cdots & & 1 \end{vmatrix} \qquad (8)$$

it can be shown by prediction theory techniques[3] that the linear redundancy in a given element due to the n previous elements is

$$R_n = \frac{1}{2} \log \frac{P^{(n)}}{P^{(n+1)}} \qquad (9)$$

As one can intuitively guess, it can also be shown that $R_n < R_{n+1}$ for all $n > 1$. The R of equation 6 is the limit of R_n for large n.

Kretzmer[1] has found that typical television images exhibit correlation coefficients that decay somewhat exponentially in both horizontal and vertical directions. For the moment, the vertical correlation is neglected to estimate the redundancy due to linear correlations among the elements along a line. If the correlation coefficients are assumed to decay exponentially, i.e., if

$$\rho_k = \rho^{|k|} \qquad (10)$$

then from equation 9 the following is obtained

$$R_n = R_1 = -\frac{1}{2} \log (1-\rho^2) \qquad (11)$$

for all $n > 1$. Thus, for an exponential correlation, the total linear redundancy is equal to the previous-element redundancy. The previous-element redundancy ranges from 0.6 to 2.83 bits per element for previous-element correlations of 0.75 and 0.99, respectively. It should be noted that for a 64-level quantized signal, the source capacity is six bits per element.

It is of value to determine the linear redundancy for nonexponential correlation decays in order to get an idea of how much of the total linear redundancy along a line is contained in the previous element. Two simple examples are shown in Figs. 1(A) and (B). Fig. 1(A) has correlation decaying linearly to zero and thereafter remaining zero, while Fig. 1(B) has constant correlation ρ for all $k > 1$. If the limit of equation 9 is taken for large n, both of these examples are found to have

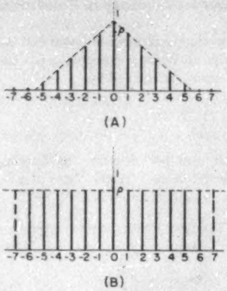


Fig. 1. Typical nonexponential correlation decays

a linear redundancy given in the equation

$$R=-\frac{1}{2}\log(1-\rho) \tag{12}$$

where ρ is the previous-element correlation. On comparing equations 11 and 12, it is seen that the previous-element redundancy differs from the total linear redundancy by 1/2 log $(1+\rho)$, which is less than 1/2 log 2, or 1/2 bit per element. Thus, for these two examples at least, a significant part of the total linear redundancy is contained in the previous element.

To demonstrate more effectively the significance of the previous-element redundancy, consider the 2-dimensional correlations in a typical picture. The small circles in Fig. 2 represent the picture elements along successive scanning lines of a television raster. The element marked X represents the instantaneous present position of the scanning spot, while the past elements are ordered according to their spatial distance from the present element. For simplicity, interlace was neglected and the line-to-line spacing was assumed equal to the distance between successive elements in a line. From Kretzmer's work, it is reasonable to assume that the correlation between any two elements is an exponential whose exponent is proportional to the spatial distance between the elements. Thus, lines of constant correlation become circles in the plane. If ρ denotes the correlation between the present element and either of elements *1* and *2*, then the correlation associated with elements *3* and *4* becomes $\rho^{\sqrt{2}}$, etc.

Fig. 3 shows the increase in linear redundancy as successively more past elements are taken into consideration. The redundancy is plotted for two typical values of adjacent-element correlation. The ordinate at $n=1$ is the previous-element redundancy, while $n=2$ gives the redundancy due to the previous element and the adjacent element of the previous line. The ordinate at $n=19$ is the redundancy in the 19 elements closest to the present element. It is seen that the redundancy levels off very rapidly so that the previous-element redundancy is a very significant portion of the total linear redundancy. The linear redundancy can be expected to lie between one and three bits for a typical television image, depending upon its complexity.

It is clear that if the frame-to-frame correlation were taken into consideration assuming an exponentially decaying correlation in 3-dimensional space, the redundancy values would not be significantly different. In fact, it has been verified that the correlation between corresponding elements on successive frames of a moving image is even less on the average than the previous-element correlation. If a coding scheme is devised to utilize the previous-element linear redundancy only, these results suggest that the incorporation of higher order linear redundancies would yield rapidly diminishing returns, probably not enough to justify the greatly increased complexity of the coding scheme. This conclusion is in complete accord with the results of Harrison,[4] who showed that the gain in signal-to-noise ratio provided by slope and planar prediction is only slightly greater than that afforded by previous-value prediction.

Nonlinear Redundancies

The preceding section was concerned with the statistical redundancy in a gaussian signal or, equivalently, the linear redundancy in a more general signal. It was shown that the total linear redundancy is relatively insensitive to variations in the form of the correlation decay, but depends mostly upon the correlation value associated with the most highly correlated element. In typical television images, that element is the previous one in time.

It is of interest to determine how much of the total statistical redundancy is represented by the linear redundancy in a nongaussian signal. Although general computations cannot be made, some hint can be obtained by considering specific examples. One such example is the case of an exponentially decaying joint probability distribution suggested by Oliver.[5]

Assume that the television signal is quantized to r levels, and that the joint probability distribution of two adjacent elements is of the form

$$p(i,j)=K\alpha^{-|i-j|},\ 1\leq i,j\leq r \tag{13}$$

where α is a constant depending upon the correlation, and K is the normalizing constant. This distribution is not an unrealistic one for television images and has the property that the most probable level for any element is the level of the previous one.

The first-order marginal distribution is given by

$$p(i)=\frac{K}{\alpha-1}(1+\alpha-\alpha^{-i+1}-\alpha^{i-r}) \tag{14}$$

This distribution is shown in Fig. 4 for $r=10$ and for values $\alpha=2.3$ and $\alpha=7.8$, corresponding to previous-element correlations of 0.87 and 0.98, respectively. It is seen that the first-order distribution is very nearly uniform, hence the nonpredictive redundancy is sufficiently small to be neglected.

The predictive redundancy in the previous element, i.e., the average information about the value of any element provided by the value of the previous element is given by (see equation 3 and the discussion there)

$$R=\sum_{i,j}^{r}p(i,j)\log\frac{p_j(i)}{p(i)} \tag{15}$$

Values of this redundancy for $r=10$ have been found to be 1.1 and 2.3 bits per sample for α values of 2.3 and 7.8, respectively. By comparison, equation 11 yields previous-element linear redundancies of 1.02 and 2.33 bits, for correlations of 0.87 and 0.98, respectively. Thus, for this example, the previous-element linear redundancy provides a fair estimate of the true previous-element redundancy even though the joint distribution is neither continuous nor gaussian.

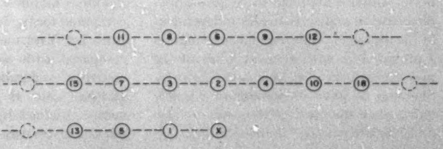


Fig. 2. Elements in a television scanning raster

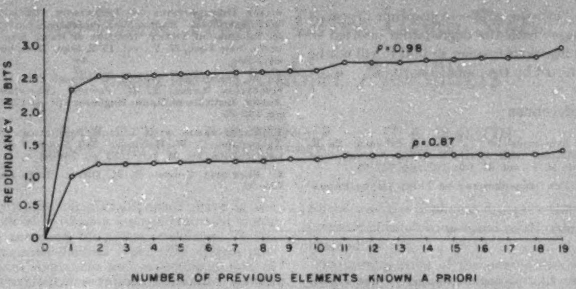


Fig. 3. Linear redundancy in a television image

A second example for comparison is afforded in the work of Schreiber,[3] who measured statistical distributions up to third order on typical television images quantized to 64 levels. Although Schreiber did not tabulate redundancies specifically, they may be obtained readily from his results. In his notation, y represents the present element and x the previous one. The total previous-element redundancy is given by $C - H_x(y)$, of which $C - H(x)$ is the nonpredictive, and $H(x) - H_x(y)$ is the predictive part. The source capacity C of a 64-level picture is, of course, six bits. His subject A had a previous-element redundancy of 2.64 bits, of which 0.3 bit is nonpredictive and 2.34 bits are predictive. His subject B had a previous-element redundancy of 4.15 bits of which 1.70 bits are nonpredictive and 2.45 bits are predictive. Thus, the previous-element predictive redundancies of both his subjects are comparable to the estimates given here.

Schreiber's third-order probability distribution measurements provide the predictive redundancy in the two preceding elements. For subject B, this redundancy is found to be 2.90 bits, which is only slightly greater than the previous-element predictive redundancy. Thus, additional evidence is given that the higher order redundancies converge rapidly, and that the previous-element redundancy constitutes a very significant part of the total redundancy.

Bandwidth Requirements

So far the discussion has concerned television redundancy but nothing has been mentioned of the minimum bandwidth required for the transmission of television images. If the television message is assumed to be disturbed by an additive white gaussian noise, 40 db below the message power, then the capacity C of the 4-mc television channel is 6.64 bits per sample element or about 53 million bits per second. If, for the moment, a gaussian message is assumed, and I denotes the information in each element of message plus noise when the infinite past history is known, then $(C - I)$ is the redundancy of the message-plus-noise signal. The redundancy R of the message only is greater than or equal to $(C - I)$, with the equality holding in the limit as the noise vanishes. Both C and I become infinite in the limit, but their difference remains finite. If a statistical coding scheme is devised so that the message can be transmitted over a second channel with bandwidth corresponding to a capacity I, the effective bandwidth reduction factor is C/I; but the relation

$$\frac{C}{I} < \frac{C}{C-R}$$

provides an upper bound to the reduction factor. Since the right-hand side becomes unity in the limit as the noise vanishes, some degradation must accompany any reduction by statistical coding.

If the coding scheme is to be stationary in time and is to operate identically on all possible images, then only the predictive part of the statistical redundancy can be effective in bandwidth compression. Although individual images may have nonpredictive redundancies due to their nonuniform distribution of brightness, the distribution over a large number of successive frames in a television program is quite uniform. The nonpredictive redundancy is quite small, if not zero.

A predictive redundancy of three bits in a television message allows a potential halving of the bandwidth of a 6-bit channel. Although this figure might seem low for the apparently high correlations that are noted in television pictures, it must be remembered that the relation between redundancy and statistical correlations is a logarithmic one.

It is often pointed out with a "big numbers" argument that the present television system is capable of handling a vast number of possible pictures most of which are random-noise patterns. It is indeed true that a 4-mc channel with 40-db signal-to-raise ratio is capable of transmitting a 100-level picture of some 200,000 elements or a total of $10^{400,000}$ possible pictures. It is also true that "most" of these pictures are utter nonsense and do not represent useful pictures. If there were evidence (and there is not) that some 99.99...9% of these pictures could be discarded, a bandwidth saving could be realized. To fix numbers, suppose that the sequence of nine's in the foregoing percentage contains 200,000 successive nine's, a large percentage indeed. To get a feeling for just how large, this percentage represents one sensible picture for every $10^{200,000}$ nonsensical ones, or if a new picture were displayed successively every 1/30 second on a television screen, one would expect a sensible picture to occur once in $10^{199,989}$ centuries. But even this large percentage still leaves a total of $10^{200,000}$ sensible pictures, and the potential bandwidth reduction factor is two. The potency of the logarithm is clearly indicated.

Although the values of statistical redundancies given in this paper do not give promise of large-scale bandwidth compression factors, it is not intended to imply that the present large bandwidth is necessary. For example, a standard television signal passed through a well-designed 2-mc linear-phase low-pass filter produces a surprisingly high-quality image. This very simple bandwidth compressor gives a 2-to-1 factor with very

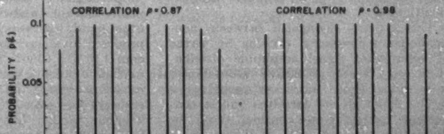


Fig. 4. First-order distribution of levels

little degradation of all pictures except the critical test pattern. The same reduction factor would be very difficult to obtain from statistical coding, which has its own share of degradation. If large compression factors are to be realized they must come, not from statistical redundancy, but from the nonstatistical redundancies mentioned in the first paragraph. One must devise intelligent ways to "fool the eye"—to transmit degraded images with the degradation inserted in an ingenious manner so that it will not be detected by the human observer.

References

1. Statistics of Television Signals, **E. R. Kretzmer.** *Bell System Technical Journal*, New York, N. Y., vol. 31, July 1952, pp. 751–63.

2. The Measurement of Third Order Probability Distributions of Television Signals, **W. F. Schreiber.** *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, New York, N. Y., vol. IT-2, Sept. 1956, pp. 94–105.

3. A Prediction Theory Approach to Information Rates, **K. H. Powers.** *Convention Record*, Institute of Radio Engineers, pt. 4, 1956, pp. 132–39.

4. Experiments with Linear Prediction in Television, **C. W. Harrison.** *Bell System Technical Journal*, vol. 31, July 1952, pp. 764–83.

5. Efficient Coding, **B. M. Oliver.** *Ibid.*, pp. 724–50.

◆

Discussion

E. Colin Cherry (University of London, London, England): As one who has been engaged upon the practical problem of television and facsimile compression for some years, I submit a few comments. For brevity, I ask that my appreciation of the paper be taken for granted, leaving me free to concentrate on the adding of emphasis to certain points which I feel have received too little attention in this paper.

Some of us in Britain have been a little impatient of the numerous published statistical analyses of television pictures and of estimations of the degree of information redundancy. Truly, such matters must be considered and figures obtained, in order to estimate the likely reward. But, having got the actual figures, what are we going to do about it? By what practical methods can we compress bandwidth? By what coding techniques? From this point of view we must admit that information theory gives us no help; it can tell us how much information is involved, or how efficient it is, but it does not purport to tell us what to do to improve our systems, nor can it.

To some people, it may seem heretical to suggest that the information measure (entropy) is not necessarily the most adequate measure, or the most appropriate, when it comes to considering presentation of signals, as in the present instance.

G. G. Gouriet, of the British Broadcasting Corporation, and I once preferred another (also possibly inadequate) measure which involved the integrated curvilinear length of the picture signal waveform.[1] Dr. Gabor subsequently showed that our measure and the information (entropic) measure give identical values for one particular distribution $P_j(i)=\exp-k^2\ (j-i)$. There are possibly many other measures but my point is that such measures might be chosen because they are directly interpretable as practical compression techniques. Perhaps we should not worship always too fervently at the shrine of information theory!

The authors distinguish between nonpredictive redundancy, (arising by virtue of nonuniform signal amplitude distributions $p(j)$, and predictive redundancy, from nonuniform transition probabilities $p_j(i)$. Although we ourselves find it convenient sometimes to consider these individually, they are not, of course, independently controllable; for it is axiomatic that if all $p_i(j)$ are made equal, by coding, then all $p(i)$ automatically become equal. Thus if all predictive redundancy is removed, the nonpredictive redundancy vanishes with it; but the converse is not true. Certainly, with stores of practical capacity it can only be the $p_i(j)$ on which we hope to operate.

Finally, I completely agree with the authors that "the best way of achieving bandwidth reduction...is by cleverly degrading the *information* in a picture in such a way that the human observer would not notice it appreciably," with the exception of the word I italicize. Entropic information rate can only have magnitude, not quality; it cannot be "degraded." Indeed we must approach the problem as a study, not only of relevant signal measures, but of their significance to human viewers.

Reference

1. Some Possibilities for the Compression of Television Signals by Recoding, **E. C. Cherry, G. G. Gouriet.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 100, pt. III, Jan. 1953, p. 9.

K. H. Powers and **H. Staras:** We must agree with Dr. Cherry that information theory does not provide us with the necessary coding schemes to maximize the efficiency of our transmission systems. However it can and does give more insight into the reasons that present systems are inefficient and provides us with the yardstick to let us know how much improvement we can expect to gain.

The use of Shannon's measure of information can hardly be regarded as an act of worship at the shrine of information theory; that measure is by no means arbitrary. It is a carefully chosen measure that satisfies certain requirements dictated by our intuition about what information really is. Among these requirements is an additivity of the amount of information from independent sources and, more important, an invariance of the information content under bi-unique transformations, i.e., 1-to-1 coding and decoding schemes. It is precisely this measure that leads to a direct proportionality between information capacity and the bandwidth utilized. On the other hand, the concept of "picture detail" proposed by Cherry and Gouriet (see reference 1 of Dr. Cherry's discussion) gives its maximum possible value to a periodic signal requiring virtually no bandwidth at all, yet registers only some 30% of its maximum for a purely random band-limited white noise for which no bandwidth compression is possible. It is difficult to see how one could regard such a measure as indicative of the ultimate bandwidth compression that can be realized without significant degradation.

It is indeed true that Shannon's measure is somewhat difficult to evaluate in the case of information displayed to the human observer, as in television.[2] That handicap should not be attributed to any inadequacy of the measure, however, but rather to the complexity of the apparatus involved. If we knew, for example, the conditional probability distribution of the true luminance of a scene for a given stimulus acting on the eye of the observer, Shannon's measure is completely adequate to compute the amount of information displayed. That measure is the most appropriate, in fact, for estimating the amount of bandwidth compression to be realized from statistical coding, as was the sole purpose of the paper under discussion.

We must agree with Dr. Cherry that the predictive and nonpredictive redundancies in a signal are not independently controllable but we cannot agree that removal of the former implies removal of the latter. The absence of predictive redundancy does not imply uniform transition probabilities, as Dr. Cherry apparently infers from our paper; it implies only that the transition probability does not depend upon the present state. Equation 3 clearly shows that when the predictive redundancy is zero, every $p_{s_j}(i)$ is equal to $p(i)$ but the $p_{s_j}(i)$ are not necessarily uniform. Our reason for separating these redundancies was not because we can perform operations on only the transition probabilities but rather because it is only the predictive part of the statistical redundancy that is significantly different from zero in a television signal.

We are happy that Dr. Cherry agrees with our conclusions that if any large-scale redundancy exists in a television system, it is of the type that we called nonstatistical (perhaps a better term would have been "physiological") redundancy. This redundancy is removed by degrading the signal specification, thereby reducing the information content in a television image. Our casual use of the word "information" in the synopsis should hardly be interpreted in the entropic sense.

Since writing the paper, the authors have been informed by Dr. Schreiber that a recent examination of stores of film clips have led him to the opinion that the frame-to-frame correlation on even fast-action motion pictures is substantially higher than either element-to-element or line-to-line correlation. Perhaps a more detailed study will lead to more optimism in the consideration of statistical coding.

Two-Phase Motor Drive Using Static Frequency-Doubler Amplifiers

G. W. DICK
ASSOCIATE MEMBER AIEE

P. P. BIRINGER
ASSOCIATE MEMBER AIEE

G. R. SLEMON
MEMBER AIEE

Synopsis: A variable-speed drive is described in which a 2-phase induction motor is energized by power modulators of the frequency-doubler type. The system provides continuous and reversible control of motor torque. In comparison with similar drives using conventional magnetic amplifiers this system has the advantages of simplicity, high efficiency, inherent zero stability, and reasonably flat torque-speed characteristics. Both steady-state and transient characteristics of the drive are described and analyzed.

FOR SOME TIME there has been a major research effort to produce speed and position control systems using a-c motors. Such systems have the advantages over d-c motor systems of using a simpler, commutatorless machine and of not requiring conversion from a-c to d-c power.

One system which has been applied extensively, especially in the low-power range, is based on the use of magnetic amplifiers as power modulators.[1,2] These amplifiers control the voltage on one phase of a 2-phase induction motor while the other phase is given a constant excitation. A more general use of this drive is limited by two major disadvantages. Because of the constant excitation on one phase the motor losses are high at standstill, and auxiliary cooling apparatus is usually necessary. Also the need for a reversible torque and shaft rotation requires that two balanced magnetic amplifiers be connected in push-pull on the control phase of the motor.[3] Other disadvantages are the harmonic content of the magnetic-amplifier output, the changing phase relation between the main and control excitations, the difficulty of maintaining zero stability of the push-pull amplifiers, and the nonlinear drooping characteristic of the torque-speed relationship obtained.

An investigation of the possibilities of using static frequency doublers for this application was undertaken because of three fundamental properties of this type of modulator:[4]

1. The phase of the second harmonic output of the frequency doubler changes by 180 degrees when the polarity of its direct control current is reversed.
2. With the control current of the doubler set at zero the second-harmonic output is zero. Other harmonic components of the output may be made essentially zero without delicate core matching.
3. A frequency doubler is similar to a saturable reactor in that it behaves approximately as a controlled current source within its control limits. It was felt that this property could be used to advantage in shaping the torque-speed relation of an induction motor.

After describing two arrangements of the proposed drive the paper presents experimental information on the steady-state characteristics of the system and on the time constant of the power modulators. An analysis of the system leading to a prediction of its characteristics is presented in the appendix.

Description of the Drive

Fig. 1 shows one arrangement of the drive in which a frequency doubler is used to supply each of the two motor fields which are spatially in quadrature. In order to provide for a 90-degree phase difference between the two field voltages, it is necessary to provide for a 45-degree phase difference in the fundamental-frequency supplies to the frequency doublers. These supplies are therefore obtained from appropriate tappings on a 3-phase autotransformer. The use of capacitors to obtain this phase difference is not recommended because of their effects on the characteristics of the doublers.

The output currents of the doublers are controlled by the direct currents in their control windings. The system of Fig. 1 provides for simultaneous control of both motor field currents. To obtain reversible torque the output of only one of the doublers can be allowed to reverse in phase when the polarity of the control signal is reversed. Thus a full-wave bridge rectifier is inserted in the control circuit of one of the doublers.

Since the control and output circuits of the doublers are similar, inductors are inserted to inhibit the flow of second-harmonic current in the control circuits. The two control circuits are connected in parallel; if a series arrangement is used the induced-harmonic voltage in the control windings may combine at the rectifier to produce a nonlinear positive feedback current in one doubler. A typical wave form of the output current of a doubler is shown in Fig. 3(A).

Fig. 2 shows a drive in which a pair of smaller doublers are used to supply each motor field. Autotransformers on the supply system are used to provide a 90-degree phase difference between the fundamental-frequency primary voltages of the two doublers which constitute a pair. The 45-degree phase difference between the primary voltages of one pair and those of the other pair is still maintained. This 90-degree phase difference permits the output windings of each pair to be connected in an additive second-harmonic sense while the control windings are connected in the opposite cancelling sense. In addition to the advantage of having no second harmonic in the control circuit this arrangement has several other advantages:

1. The 4th, 8th, etc., harmonics are cancelled in the output circuit resulting in an approximately sinusoidal-wave form output as shown in Fig. 3(B). The lowest harmonic producing distortion is the 6th, and it can be seen from Fig. 3(C) that this component will be relatively small.
2. Only 4th, 8th, etc., harmonics remain in the control circuit. Not only are these easier to filter but they may be shorted out if desired without affecting the second-harmonic output. The control circuit inductors can therefore be eliminated. In addition to reducing the weight and cost of the modulator, a considerable improvement in transient response is obtained.

Either of the drives shown in Fig. 1 and Fig. 2 may be used with control on only one motor field while the other field is supplied with constant excitation as in conventional magnetic amplifier driven systems. For this mode of operation no rectifier is required in the control circuit.

Data on Experimental Drive Components

In the experimental drive two frequency doublers were used in the arrangement of Fig. 1. The cores of these doublers con-

---

Paper 57-631, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted October 24, 1956; made available for printing April 18, 1957.

G. W. Dick, P. P. Biringer, and G. R. Slemon are with the University of Toronto, Toronto, Ont., Canada.

The investigation reported herein was made possible through support extended to the Department of Electrical Engineering, University of Toronto, by the Defense Research Board of Canada, under Grant No. *4001-02*, Project *048-40-01-02*.

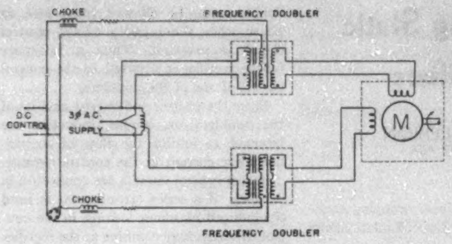


Fig. 1. Connections for basic induction-motor drive using two frequency doublers

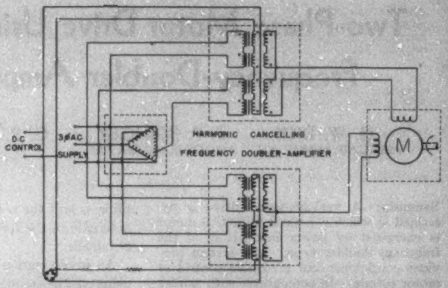


Fig. 2. Connections for induction-motor drive using four frequency doublers

sisted of standard 0.014-inch silicon iron laminations of the *E* and *I* types. The over-all core dimensions were 4.5 inch by 3.75 inch with a stacked thickness of 2 inch. The secondary winding (252 turns) and control winding (215 turns) were placed on the center leg while the primary windings with a total of 430 turns were placed on the outer legs. The supply voltage to the doubler primary winding was 115 volts, 60 cycles.

The induction motor was a Diehl, 2-phase servomotor Type *S.S.FPF66-11-1* rated at 50 watts output for a 115-volt 60-cycle supply. Fig. 4 shows torque-speed curves obtained for this motor when tested with a variable 2-phase 120-cycle supply voltage.

Steady-State Characteristics

Fig. 5 shows the torque-speed characteristics of the drive of Fig. 1 for various values of doubler control current. With reversed polarity of the control current the characteristics of the drive are identical to those shown but with reversed direction of rotation.

It will be noted that at any value of control current the torque is reasonably constant over most of the speed range. The shape of the torque-speed characteristic differs from that obtained with constant voltage excitation in Fig. 4 because of the approximate current-source property of the doublers. If an ideal current source were used to supply an induction motor, in which the magnetizing current could be neglected, the torque would vary as the inverse of the motor slip. The analysis of the doubler in Appendix I shows that the doubler current is not constant but decreases as the load impedance is increased; at high load impedances the doubler becomes essentially a voltage source. The magnitude of this decrease in output current with load is adjustable to a considerable extent in the design of the doublers. The doubler characteristic may therefore be shaped in such a way as to counteract the tendency of the torque to rise with speed at constant current. This would result in a torque which would remain approximately constant up to the region of synchronous speed where the torque would fall sharply toward zero due to the high motor impedance.

The torque-speed characteristics of Fig. 5 were obtained with a doubler which was not specially designed for the motor. The reduction in output current as the motor impedance increases is not sufficiently large to prevent a gradual rise in torque over most of speed range. No attempt has been made as yet to determine the optimum parameters of a motor for use with this system. It is possible that a machine having a considerably lower ratio of rotor resistance to rotor reactance could have been used to advantage.

Linearity of Control

With excitation from a voltage source the torque of a 2-phase induction motor is proportional to the voltage on each phase. For this reason, when power modulators having a linear relation between control signal and output are used, only one phase of the motor is controlled. With control on both phases the torque would be expected to increase as the square of the control signal.

In the analysis of Appendix I it is shown that the output current of the doubler is a nonlinear function of the control current of the usual saturation curve shape. This nonlinearity tends to compensate for the square-law relationship between torque and motor current resulting in the approximately linear relationship between torque and control current for control on both phases. See Fig. 5.

The torque-speed characteristics of a drive with control on only one field are shown in Fig. 6. It is noted that the relationship between torque and control current is of a nonlinear saturating type. This nonlinearity is due both to the nonlinearity in the output current-control current relationship and to a shift in the phase difference between the two field currents.

Efficiency

Fig. 7 shows curves of the efficiency of a drive using one doubler per phase. The losses in the drive are mainly in the motor. The doubler losses are largely core losses which are independent of the control signal level; the copper losses are made considerably lower than the core losses to minimize saturated reactance effects in the doubler. The efficiency of the doubler at maximum power output would be of the order of 70 to 95% dependent on size and frequency.[4] The arrangement of Fig. 2 with two doublers per field would have an efficiency somewhat greater than that shown in Fig. 7 since losses in the motor windings due to higher harmonic currents would be largely eliminated.

Time Constant of Doubler

The transient response of the doubler to a step function of control voltage was found to be a single exponential function. With the recording apparatus available it was impossible to detect any time delay preceding the exponential rise of second-harmonic current. The incremental time constant was measured over a range of control-current levels and plotted in Fig. 8. The time constant of the control circuit with no alternating voltage present is also drawn in Fig. 8 showing that the time constant of the doubler is essentially that of its control circuit. In order to

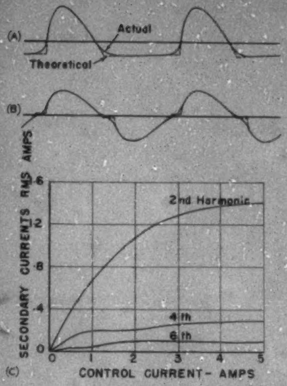


Fig. 3. Output current of frequency doubler with R-L load

A—Single doubler at 30% maximum output current

B—Pair of phase-shifted doublers at 30% maximum output current

C—Harmonic components of output current

obtain measurable time constants with the apparatus available a very low value of control circuit resistance was used. Increase of the control circuit resistance would result in a proportional decrease in the time constant until the time constant became of the same order as any time delay present.

The analysis of Appendix II shows the effect of control current and load impedance on the effective inductance of the control circuit. The rapid decrease of the time constant from its maximum value at zero control current to a value dependent on the saturated inductance is indicated both by theoretical and experimental results.

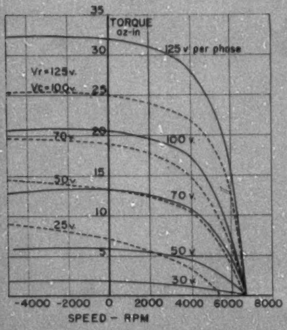


Fig. 4. Torque-speed characteristics of 2-phase servomotor on sinusoidal, 120-cycle voltage source

- - - - One phase at 125 volts; various voltages applied to other phase

——— Various equal voltages on both phases

The time constant of a doubler drive would be increased considerably if an inductor were inserted into the control circuit to control second-harmonic currents. Therefore where speed of doubler response is of importance the arrangement of Fig. 2 with two doublers per field and no control circuit inductors is preferable. This arrangement also eliminates any mutual inductance between the control and output circuits.

Comparison with Magnetic Amplifier Drive

The efficiencies of doublers are comparable with those of transformers of similar size. Further gains in efficiency for a doubler-driven system may be achieved by a redesign of the 2-phase motor.

In conventional magnetic-amplifier drives where the excitation on one field is maintained constant, a separate fan motor is usually required to provide cooling when the motor is at or near standstill. When both fields are controlled simultaneously the motor losses at standstill and zero torque are zero, and accordingly the design of self-cooled motors becomes feasible.

In a magnetic amplifier system, reasonably well regulated power and bias supplies are required to maintain zero output at zero control signal. With a doubler system, only an approximate matching of the cores is required to obtain a stable zero output with zero control signal; no fixed bias is required. Also the output current of the doubler displays the same relative independence to changes in supply voltage as it does to changes in load impedance. Line-voltage fluctuations are therefore unimportant except near the maximum control current level.

The dynamic power gain of a doubler is at least one order of magnitude less than that of push-pull magnetic amplifiers of comparable rating. The magnetic amplifier achieves its high gain by the use of feedback. The doubler is more nearly comparable with a saturable-reactor system in this application since it has no feedback. In a complete control system the deficiency in power gain of the doubler may be made up by the use of appropriate preamplifiers.

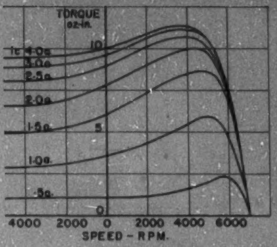


Fig. 5. Torque-speed characteristics of basic frequency-doubler drive

Control on both phases

Maximum doubler voltage at standstill = 75 volts

Conclusions

The investigations which were made of the feasibility of using static frequency doublers as power modulators for a 2-phase induction motor drive have shown that the system has several desirable properties:

1. The relationship between motor torque and control signal can be made reasonably linear up to the saturation level of the doublers.
2. The motor torque at any value of control signal can be made more or less independent of motor speed in the forward or reverse direction except in the region near synchronous speed where the torque drops to zero. In a control system the action of this drive would be similar to that of a d-c motor with constant armature current and control of the field current.
3. The system has inherent current, torque, and speed limits.
4. The system has inherent zero stability.

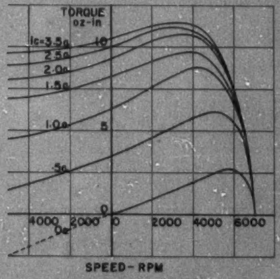


Fig. 6. Torque-speed characteristics of basic frequency-doubler drive

Control on one phase only

Control current in doubler supplying other phase: $I_c = 3.5$ amp

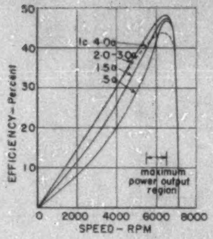


Fig. 7. Efficiency-speed characteristics of basic frequency-doubler drive for various values of control current (I_c)

Maximum motor efficiency with 125-volt 60-cycle sinusoidal voltage supply=61%

5. A high over-all efficiency can be achieved with the drive because the motor is operated with a balanced supply and the doubler efficiency is inherently high.
6. Since the motor losses are zero at standstill self-cooling of the motor is feasible.
7. The response of the system is reasonably independent of supply voltage variations and no bias supply is required.
8. Ordinary silicon iron can be used in the frequency doublers although some improvement in response can be obtained by the use of materials having a lower saturated permeability.

The work reported in this paper was limited to showing the feasibility of the proposed drive. The doublers used in the tests were not designed for this purpose but were those already available for student demonstration purposes. Investigations in the optimum design of induction motors and doublers for this type of drive, in the use of feedback to improve the power gain of the doublers, and in the properties of this drive when used in a position or speed control system are being continued.

Nomenclature

E_m=peak value of line voltage, volts
$N_{p,s,c}$=number of turns in primary, secondary, control winding
$L_{p,s,c}$=unsaturated inductance of primary, secondary, control winding
I_k=knee-point magnetizing current in primary winding
$i_{p,s(t)}$=instantaneous value of primary or secondary current
k=ratio of saturated permeability to unsaturated permeability
L_L=external secondary load inductance, henries
R_L=external secondary load resistance, ohms
$L_a=L_L+4\,\frac{k}{1+k}\,L_s$ henries
$Z_a=\sqrt{R_L{}^2+\omega^2L_a{}^2}$ ohms
$\theta=\tan^{-1}\frac{\omega L_a}{R_L}$ radians
$\phi_{1(t)}$=flux in core no. 1, lines per square centimeter (cm^2)
$I_{s,p(x)}$=initial values of currents at switching time x
α,β=switching angles in radians, $\alpha=\omega a$
$n=\frac{R_L}{\omega(L_L+2L_s)}\ (\pi+\alpha-\beta)$
$m=\frac{R_L}{\omega L_L}\ (\beta-\alpha)$
a,b=switching times, seconds
I_s=2nd harmonic output current, root mean square amperes
A=net area of one core, cm^2
l=effective flux path length, centimeters
μ=unsaturated core permeability centimeter-gram-second (cgs) units
$\phi_{avg.}$=the average flux in one core over one complete line voltage cycle
L_{cn}=effective inductance of the control circuit which determines the doubler incremental time constant, henries
ϕ_s=saturation or knee-point flux density
L_{xs}=saturated inductance of winding x

Appendix I

Steady-State Transfer Function of Frequency Doubler

In the analysis the following assumptions will be made:

1. Hysteresis is neglected.
2. The B-H loop is approximated by two straight lines giving a saturated permeability K times the unsaturated permeability.
3. All impedances to second harmonic currents are lumped in the secondary circuit and the control circuit impedance is assumed to be infinite.
4. The cores are identical.
5. Primary and secondary winding resistances are negligible.
6. The design is such that at no load each core is carried from positive to negative saturation.
7. The leakage fluxes are negligible in comparison with the mutual flux in each core.

The circuit of a frequency doubler is shown in Fig. 9. The unsaturated inductances of the primary, secondary, and control windings are L_p, L_s, and L_c respectively. The saturated values are indicated by use of an additional subscript s.

At some time $t=a$ during the negative half cycle of the supply voltage, core 2 reaches negative saturation. At this instant the primary, secondary, and control winding currents are $I_{p(a)}$, $I_{s(a)}$, and $I_{c(a)}$. These initial conditions are related to the knee-point magnetizing current I_k of core 2 by the following equation

$$\sum NI_{(a)}=N_pI_{p(a)}-N_sI_{s(a)}-N_cI_c=-N_pI_k \tag{1}$$

At $t=a$ the core permeability changes from μ to its saturated value $\mu_s=k\mu$. The differential equations of the system are

$$E_m\sin(\omega\tau+\alpha)=(L_p+L_{ps})\frac{di_p(\tau)}{d\tau}+(M_{ps}-M_{pss})\frac{di_s(\tau)}{d\tau} \tag{2}$$

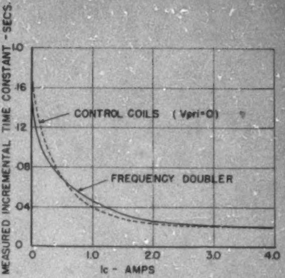


Fig. 8. Time constant-control current curves for frequency doubler

---- Measured incremental $4L_{cn}$

$$(L_s+L_{ss})\frac{di_s(\tau)}{d\tau}+(M_{ps}-M_{pss})\frac{di_p(\tau)}{d\tau}=-L_L\frac{di_s(\tau)}{d\tau}-R_Li_{s(\tau)} \tag{3}$$

where the time scale is translated by $\tau=t-a$, and $\omega t=\omega\tau+\alpha$. Assumption 7 is used to reduce the number of pertinent inductances. Solving these equations by the Laplace transformation gives the following solution in the real time domain:

$$\underset{a<t\leq b}{i_{s(t)}}=-\frac{N_s}{N_p}\frac{1-k}{1+k}\frac{E_m}{Z_a}\sin(\omega t-\theta)-\left[\frac{N_s}{N_p}\frac{1-k}{1+k}\frac{E_m}{Z_a}\sin(\theta-\alpha)-I_{s(a)}\right]\times\epsilon^{-\frac{R_L}{L_a}(t-a)} \tag{4}$$

$$\underset{a<t\leq b}{i_{p(t)}}=\frac{N_c}{N_p}I_c-I_k+\frac{N_s}{N_p}\frac{1-k}{1+k}\left[I_{s(a)}\frac{2}{1-k}-\epsilon^{-\frac{R_L}{L_a}(t-a)}\right]+\left(\frac{N_s}{N_p}\frac{1-k^2}{1+k}\right)\frac{E_m}{Z_a}\left[\sin(\omega t-\theta)+\sin(\theta-\alpha)\epsilon^{-\frac{R_L}{L_a}(t-a)}\right]+\frac{E_m}{\omega L_p(1+k)}(\cos\alpha-\cos\omega t) \tag{5}$$

where

$L_a=L_L+\frac{4L_{ss}}{1+k}$

$Z_a=R_L+j\omega L_a$

$\theta=\tan^{-1}\frac{\omega L_a}{R_L}$

and b=time at which core 2 re-enters its unsaturated region.

The time at which the total flux of the two cores reaches its maximum negative value is determined to obtain an expression for $I_{s(a)}$. The flux in core 1 is

$$\phi_{1(t)}=0.4\pi\mu\frac{A}{l}(N_pi_{p(t)}+N_si_{s(t)}+N_cI_c)$$

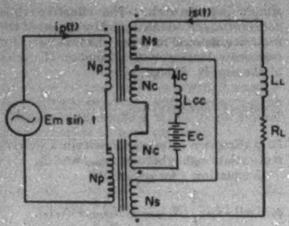


Fig. 9. Static frequency doubler

and in core 2

$$\phi_{2(t)}=0.4\pi\mu k\frac{A}{l}(N_p i_{p(t)}-N_s i_{s(t)}-N_c I_c+N_p I_k)-0.4\pi\mu\frac{A}{l}N_p I_k$$

$$\phi_{1(t)}+\phi_{2(t)}=0.4\pi\mu\frac{A}{l}[(N_p(1+k)i_{p(t)}+N_s(1-k)i_{s(t)}+N_c(1-k)I_c-N_p(1-k)I_k)] \qquad (6)$$

From equation 2 it can be seen that at $t=0$

$$\frac{d(\phi_{1(t)}+\phi_{2(t)})}{dt}=0$$

assuming zero supply impedance and zero resistance in the winding. The minimum total flux is determined by the supply voltage and, using assumption 6, is given by

$$(\phi_1+\phi_2)_{\min}=-0.4\pi\mu\frac{A}{l}(2N_p I_k) \qquad (7)$$

Substituting equation 7 into equation 6 at $t=0$ gives

$$N_p(1+k)i_{p(o)}-N_s(1-k)i_{s(o)}+N_c I_c(1-k)+N_p I_k(1+k)=0 \qquad (8)$$

and substituting equations 4 and 5 into equation 8 gives

$$N_c I_c+N_s I_{s(a)}+N_p I_k(\cos\alpha-1)=0 \qquad (9)$$

This equation applies for all values of control and load currents. Since I_k is small it is possible in many power modulator applications to use the approximation

$$N_c I_c=-N_s I_{s(a)}$$

At some time $t=b$ with a corresponding angle $\omega t=\beta$ core 2 will again re-enter its unsaturated region. At this instant

$$\sum NI_{(\beta)}=N_p i_{p(\beta)}-N_s i_{s(\beta)}-N_c I_c=-N_p I_k \qquad (10)$$

Substituting equations 4 and 5 into 10 gives the secondary current $I_{s(a)}$ in terms of the switching angles α and β.

$$N_s I_{s(a)}=\frac{-N_p I_k(\cos\alpha-\cos\beta)+\dfrac{N_s^2}{N_p}\dfrac{1-k}{1+k}\dfrac{E_m}{Z_a}\times\left[\sin(\beta-\theta)+\sin(\theta-\alpha)\epsilon^{-\frac{R_L}{\omega L_a}(\beta-\alpha)}\right]}{1-\epsilon^{-\frac{R_L}{\omega L_a}(\beta-\alpha)}} \qquad (11)$$

It is now necessary to define the relation between α and β for steady-state conditions and to determine the current during the period when both cores are unsaturated. From symmetry, core 1 will reach its saturation point at $\omega t=\alpha+\pi$. During the period $\beta<\omega t\leq\alpha+\pi$ both cores are unsaturated and circuit conditions are described by

$$E_m\sin\omega t=2L_p\frac{d_{ip(t)}}{dt} \qquad (12)$$

and

$$R_L i_{s(t)}+(L_L+2L_s)\frac{d_{is(t)}}{dt}=0 \qquad (13)$$

Solution of these equations gives for the period $\beta<\omega t\leq\alpha+\pi$

$$i_{s(t)}=I_{s(b)}\epsilon^{-\frac{R_L(\omega t-\beta)}{\omega(L_L+2L_s)}} \qquad (14)$$

$$i_{p(t)}=I_k(\cos\beta-\cos\omega t)-I_{p(b)} \qquad (15)$$

To obtain a matching equation in terms of α and β equation 14 can be stated for $t=(a+\pi/\omega)$ using the symmetry relation, $i_s(\pi/\omega+a)=i_{s(a)}$ giving

$$I_{s(a)}=I_{s(b)}\epsilon^{-n} \qquad (16)$$

where $n=\dfrac{R_L(\pi+\alpha-\beta)}{\omega(L_L+2L_s)}$

and stating equation 4 for $t=b$ giving

$$I_{s(b)}=-\frac{N_s}{N_p}\frac{1-k}{1+k}\frac{E_m}{Z_a}\sin(\beta-\theta)-\left[\frac{N_s}{N_p}\frac{1-k}{1+k}\frac{E_m}{Z_a}\sin(\theta+\alpha)-I_{s(a)}\right]\epsilon^{-m} \qquad (17)$$

where $m=\dfrac{R_L}{\omega L_L}(\beta-\alpha)$

The three equations 10, 16, and 17 can now be combined to give

$$(\cos\beta-\cos\alpha)=\frac{2\omega L_s}{Z_a}\frac{1-k}{1+k}\frac{(1-\epsilon^{-n})}{(1-\epsilon^{-n-m})}[\sin(\beta-\theta)+\sin(\theta-\alpha)\epsilon^{-m}] \qquad (18)$$

The equations 9, 11, and 18 define the doubler operation for any value of control current and may be solved numerically for sets of values of α β and I_c for any given load impedance. A typical wave form of secondary current is shown in Fig. 10.

For no load operation Z_a is infinite and from equation 18 it is seen that $\alpha=-\beta$. The average secondary voltage for this condition is

$$E_{s(avg)}=\frac{2}{\pi}\frac{N_s}{N_p}\frac{1-k}{1+k}E_m(1-\cos\alpha) \qquad (19)$$

For a purely inductive load equation 18 is indeterminate and to solve for α and β it is necessary to expand the exponential terms into power series before letting R_L approach zero. This gives the relation

$$(\cos\beta-\cos\alpha)\left[1+\frac{2L_s}{L_a}\frac{1-k}{1+k}\frac{\dfrac{\pi+\alpha-\beta}{L_L+2L_s}}{\dfrac{\pi+\alpha-\beta}{L_L+2L_s}+\dfrac{\beta-\alpha}{L_L}}\right]=0$$

where it is again seen that $\alpha=-\beta$. Equation 11 is also indeterminate but $I_{s(a)}$ may be found from the following equation.

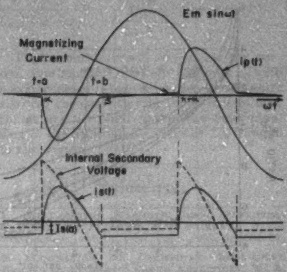


Fig. 10. Primary and secondary currents of doubler for load with $\omega L_a/R_L=0.5$

From equation 4:

$$\underset{a<t\leq b}{i_{s(t)}}=\frac{N_s}{N_p}\frac{1-k}{1+k}\frac{E_m}{\omega L_a}(\cos\omega t-\cos\alpha)+I_{s(a)}$$

From equation 14:

$$i_{s(t)}=I_{s(a)}\ b<t\leq\frac{\pi}{\omega}+a$$

Also $\displaystyle\int_a^{\frac{\pi}{\omega}+a} i_{s(t)}dt=0$

From these equations we obtain

$$I_{s(a)}=\frac{N_s}{N_p}\frac{1-k}{1+k}\frac{E_m}{L_a}\frac{2}{\pi}(\sin\alpha-\alpha\cos\alpha) \qquad (20)$$

which can be used with equation 9 to find sets of values of α and I_c.

Solution of equation 18 for general impedance loads is difficult unless some approximations are made. If the ratio $2\omega L_s/Z_L$ is assumed $\gg 30$ as is generally the case for appreciable power modulation the exponential series expansions can be truncated to give the following simplified form of equation 18:

$$\cos\beta-\cos\alpha=\frac{(\pi+\alpha-\beta)\cos\theta}{1-\epsilon^{-\frac{R_L}{\omega L_a}(\beta-\alpha)}}\times\left[\sin(\beta-\theta)+\sin(\theta-\alpha)\epsilon^{-\frac{R_L}{\omega L_a}(\beta+\alpha)}\right] \qquad (21)$$

An iterative method can be used to solve this equation.

The second harmonic component of output current may be found by Fourier analysis of equations 4 and 14. Somewhat simpler expressions are obtained if the internal secondary voltage $e_{s(t)}$ described by the equations

$$\underset{\alpha<\omega t\leq\beta}{e_{s(t)}}=\frac{N_s}{N_p}\frac{1-k}{1+k}E_m\sin\omega t$$

and

$$\underset{\beta<\omega t\leq\pi+\alpha}{e_{s(t)}}=R_L I_{s(a)} \text{ assuming } \omega(L_L+2L_s)\gg R_L$$

is analyzed giving the Fourier coefficients

$$a_1=\frac{4}{3\pi}E_m\frac{N_s}{N_p}\frac{1-k}{1+k}[(\cos\beta-\cos\alpha)-(\cos^3\beta-\cos^3\alpha)]+\frac{R_L I_{s(a)}}{\pi}(\sin 2\beta-\sin 2\alpha) \qquad (22)$$

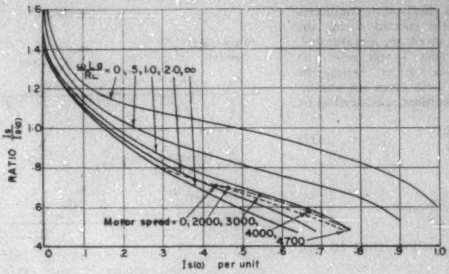


Fig. 11. Secondary-current characteristics of doubler with R-L loads

----Locus of secondary current ratio when loaded by motor at standstill and with various values of I_c

–·–·–Locus of secondary current ratio when loaded by running motor with constant control current on doublers (I_c=1.5 amp)

and

$$b_1=\frac{4}{3\pi}E_m\frac{N_s}{N_p}\frac{1-k}{1+k}(\sin^3\beta-\sin^3\alpha)+\frac{R_LI_{s(a)}}{\pi}(\cos 2\beta-\cos 2\alpha) \quad (23)$$

The root mean square value of the second harmonic current is then

$$I_s=0.707\frac{\sqrt{a_1^2+b_1^2}}{\sqrt{R_L^2+(2\omega L_a)^2}} \quad (24)$$

The variation of second harmonic current with load impedance and control current can be derived from Fig. 11. In this figure all quantities have been normalized using the base quantities:

$$E_{s(\text{base})}=E_m\frac{N_s}{N_p}\frac{1-k}{1+k}$$

and

$$Z_{s(\text{base})}=\sqrt{R_L^2+(\omega L_a)^2}$$

Where $\omega L_s/R_L \geq 30$ equation 9 is approximated by $I_c=I_{c(a)}$ in per unit. Thus the curves of Fig. 11 show approximately the variation in the ratio of second-harmonic current to control current at constant load impedance. Also if the control current is held constant the slopes of the curves show the extent to which the doubler acts as a constant current source as the load impedance is changed. The end point of each curve represents the condition of maximum secondary voltage at which point the doubler acts as a voltage source.

At variable speed the load impedance presented by an induction motor varies both in magnitude and phase. Fig. 11 includes a locus of the second harmonic current at various speeds with fixed control current and also a locus of second harmonic current for various control currents with the motor stalled.

The relationships between motor torque, speed, and control current are derived by first drawing the appropriate impedance locus on Fig. 11. Then the motor torque is found using the equivalent circuit of the induction motor and the fact that the motor torque varies as the square of its rotor current and inversely as its slip. Results calculated by this method are qualitatively in good agreement with the test values shown in Fig. 5. In the region of maximum control current the results are however less accurate because of the effects of the rounded knee-point on the **B-H** loop of the silicon-iron cores used. A more accurate approximation to the **B-H** characteristic appears to be necessary to obtain good quantitative results in this region of operation.

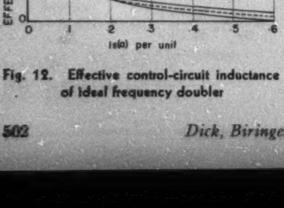


Fig. 12. Effective control-circuit inductance of ideal frequency doubler

Appendix II

Analysis of Doubler Transient Response

It is the purpose of this analysis only to predict in a qualitative fashion the variation of the incremental time constant of a basic frequency doubler operating with various circuit parameters and initially steady-state control signals. The effects on transient behaviour, of the mutual coupling between control and secondary windings, do not need to be considered here, as they can be accounted for where necessary by well known circuit theory once the basic rate of change of flux with respect to a change of control signal is known. In the circuit of Fig. 2 the mutual coupling has no effect.

The assumptions upon which this analysis are based are the same as those in appendix I except that, for simplicity, the saturated reactance of the core will be first assumed zero. The effects of saturated reactance will be considered later.

The response of the doubler to small control signal increments, is basically that of a simple lag network. The effective time constant L_{cs}/R will be found by means of the following general relation which is valid for transients lasting over a number of cycles of the supply voltage.

$$L\text{ control (effective)}=n_c\frac{d\phi_{\text{avg control}}}{dI_c} \quad (25)$$

It is therefore necessary to obtain a steady-state relationship between ϕ_{avg} and I_c.

In equation 1 for $\alpha<\omega t\leq\beta$

$$\phi_{1(t)}=0.4\pi\mu\frac{A}{l}(N_pi_{p(t)}+N_si_{s(t)}+N_cI_c)$$

Substituting from equations 4 and 5:

$$\phi_{1(t)}=0.4\pi\mu\frac{A}{l}\left[2N_cI_c-N_pI_k+2N_sI_{s(a)}+N_p\frac{E_m}{L_p\omega}(\cos\alpha-\cos\omega t)\right]$$

Substituting for $N_cI_c+N_sI_{s(a)}$ from equation 9 gives:

$$\phi_{1(t)}=0.4\pi\mu\frac{A}{l}N_pI_k(1-2\cos\omega t)$$

$$\alpha<\omega t\leq\beta=\phi_s(1-2\cos\omega t)$$

Similarly for $\beta<\omega t\leq\pi+\alpha$

$$\phi_{1(t)}=0.4\pi\mu\frac{A}{l}[2N_cI_c+N_sI_{s(\beta)}(1+e^{-n})+N_pI_k(\cos\beta-\cos\omega t-1)]$$

For $\pi+\alpha<\omega t\leq\pi+\beta$

$$\phi_{1(t)}=0.4\pi\mu\frac{A}{l}N_pI_k=\phi_s$$

For $\pi+\beta<\omega t\leq 2\pi+\alpha$

$$\phi_{1(t)}=0.4\pi\mu\frac{A}{l}[N_pI_k(\cos\beta-\cos\omega t-1)-N_sI_s(b)(1-e^{-n})]$$

$$\phi_{1\text{ avg}}=\frac{1}{2\pi}\int_{\alpha}^{2\pi+\alpha}\phi_{1(t)}d\,\omega t$$

Evaluating this integral with appropriate substitutions from equations 8 and 16 and the assumption that $\omega(L_L+2L_s)/R_L\geq 30$

$$\phi_{1\text{ avg}}=\frac{\phi_s}{\pi}\left[\pi-\sin\beta+\sin\alpha-\cos\alpha\times(\pi+\alpha-\beta)+\frac{N_sI_{s(a)}}{N_pI_k}\frac{R_L}{\omega(L_L+2L_s)}\times\frac{(\pi+\alpha-\beta)^2}{2}\right] \quad (26)$$

At $-\alpha=\beta=\pi/2$ this reduces to the expected value

$$\phi_{1\text{ avg}}=\phi_s\left(1-\frac{2}{\pi}\right)$$

From equations 8, 25, and 26 the relation for effective control circuit inductance is obtained as:

$$2L_{cs}=2L_c\frac{d}{dN_cI_c}\frac{1}{\pi}\left\{N_pI_k[\pi+\sin\alpha-\sin\beta-(\pi+\alpha-\beta)\cos\alpha]+N_sI_{s(a)}\times\frac{R_L}{\omega(L_L+2L_s)}\frac{(\pi+\alpha-\beta)^2}{2}\right\} \quad (27)$$

The differentiation indicated has been done numerically for a few values of $\omega L_L/R_L$ with the results shown on Fig. 12 in per unit of load impedance. In order to get the value of the incremental inductance at $I_c=0$ it is necessary to reduce equation 27 by means of power series expansions in terms of α and β. The relation between α and β for very small angles is also obtained in the same manner from equation 18 with the result:

$$2L_{ce}=2L_{cu}\frac{\left(1+\frac{2}{\pi}\alpha\ldots\right)}{1-\frac{2}{\pi}\frac{L_s}{L_L}(\alpha-c_1\alpha^3\ldots)} \qquad (28)$$

For $\alpha=0$ this is simply the unsaturated inductance of the control winding. For a pure resistive load,

$$2L_{ce}=\left(\frac{N_c}{N_s}\right)^2\frac{R_L}{\omega}\left[\frac{\pi}{2}+\frac{7}{4}\alpha\ldots\right] \qquad (29)$$

These two solutions seem inconsistent in that the inductance at zero control current, i.e., $\alpha=0$, for a resistive load depends on the load resistance only, while for an inductive load it depends on the unsaturated value of inductance.

Inspection of equation 28 however shows that for large values of L_s/L_L the control circuit inductance for very small values of α becomes related directly to the load impedance.

Since the effects of the saturated inductance of the cores are neglected, the practical significance of the curves of Fig. 12 is limited. Tests on a silicon-iron cored doubler (Fig. 8) show that the control-circuit time constant approaches a value determined by the saturated inductance of the core at large control levels.

This effect would be expected to be similar to the effect of secondary saturated inductance as shown in Appendix I where the term $4L_{ss}/1+k$ is added in series with Z_L. It is therefore assumed that the saturated inductance effect in the control circuit will be a constant term $4L_{cs}/1+k$ adding to equation 27 to give the following equation:

$$L_c\text{ Total}=2L_{ce}+4\frac{L_{cs}}{1+k}$$

This second term will determine the control time constant, for high values of I_c, with good accuracy; but as in the steady-state analysis the curvature in the B-H loop near the saturation point makes the determination of the appropriate value of L_{cs} difficult in this region.

References

1. An Improved Magnetic Servo Amplifier, C. W. Lufcy, A. E. Schmid, P. W. Barnhart. *AIEE Transactions*, vol. 71, pt. I, Sept. 1952, pp. 281–89.

2. Elimination of Asymmetry Zero-Drift Errors in Magnetic Servo Amplifiers, W. A. Geyger. *Ibid.*, vol. 74, pt. I, May 1955, pp. 239–45.

3. Magnetic Amplifiers for the Control of Servomotors, Allen R. Perrins. *AIEE Special Publication T-86*, "Conference on Magnetic Amplifiers," July 1956, pp. 246–54.

4. Design of the Resistively Loaded Static Frequency Doubler, P. P. Biringer. *Ibid.*, vol. 75, pt. III, Oct. 1956, pp. 834–39.

◆

Discussion

E. W. Manteuffel (General Electric Company, Ithaca, N. Y.): A new and interesting approach to the control of 2-phase induction motors has been made by the authors. While good zero stability and high efficiency seem to make frequency doublers suitable for this type of control, the lack of internal positive feedback presents another problem which needs serious consideration. In most practical applications, a push-pull magnetic amplifier will be necessary to furnish the considerable amount of control power to saturate the reactors. The response time will be increased, the over-all efficiency decreased. Also, the zero stability of the system again will be somewhat deteriorated.

Since magnetic frequency doublers require a relatively smooth d-c current for proper switching of the cores, a magnetic amplifier driving the control circuit of the doubler requires additional filtering components at its output, which would again, unfortunately, affect the over-all response time. It is understood that the statement of the authors that the motor losses at standstill are zero, applies only for the case that no torque at standstill is required.

H. Hollitscher (Canadian General Electric Company Ltd., Peterborough, Ont., Canada): The authors deserve much credit for this new approach to duo-directional control of a-c motors by using frequency doublers as power modulators instead of push-pull magnetic amplifiers. The high efficiency of frequency doublers, their zero stability, and the effect of a 180-degree phase shift in output when reversing the polarity of the control current, makes this type of power modulator very suitable for a-c motor control.

On the other hand the complete absence of any feedback to the control reactors requires a considerable amount of control power to drive the reactors into saturation, hence lowering the over-all efficiency of the drive especially for smaller units. For larger motors the control signal has to be amplified and this amplification increases the total time constant as well as reduces somewhat the efficiency of the drive. In some applications this could be a limiting factor.

Another point to be considered is the statement that the motor losses are zero at standstill. It is an inherent property of an induction motor that torque can be produced only if there are rotor losses. These are directly proportional to the torque produced. In the case of a squirrel cage motor these losses have to be transferred from the power supply by the stator field and across the air gap to the rotor. The aforementioned statement is therefore only true if no torque is applied to the shaft. By controlling one field only the motor field changes from a circular to an elliptic field. However, if both motor fields are controlled simultaneously, the motor field tends to maintain a circular shape.

It would be of interest to compare the total losses for both cases with only one phase controlled and with control in both fields, at low speed and standstill for the same motor when a certain torque is developed. It is expected that the losses in the second case will be lower but it is felt that the difference might not be big enough to omit provisions for cooling.

Andrew A. Halacsy (Jeffries Transformer Company, Los Angeles, Calif.): The paper points rightly to a disadvantage of the magnetic amplifier which is the high harmonic content of its output. The arrangement of Fig. 2 with two doublers is therefore rather interesting. The high harmonic content is, however, inherent in any device using saturable reactors for current regulation and further research in the possibility of reducing those harmonics would be welcomed. Perhaps we will be forced to think in terms of basically different regulating means in order to get rid of harmonics. This problem gains importance by the expansion of electrical controls. A comparison of Figs. 4, 5, and 6 shows that simple regulating transformers will not solve the problem.

The proper matching of motor and frequency doubler is also important. Obviously even better speed characteristics and efficiency can be achieved by this means. It would be interesting to know whether the authors have since done any further work in this respect.

G. W. Dick, P. P. Biringer, and G. R. Slemon: We wish to thank the discussers for their helpful criticisms of our paper. Mr. Hollitscher and Dr. Manteuffel have commented on the effects of a lack of internal positive feedback in the amplifiers. Although feedback can be used with the doubler amplifiers, it was not used in the application because a simple modulator with high zero-stability and duo-directional control characteristic was desired. It is true that the power gain of these amplifiers is not nearly as great as that normally obtained in conventional magnetic amplifiers. However, the objective in this system was to obtain an efficient high-power modulator using standard transformer core materials. The power gain of these amplifiers is at least as large as that of saturable reactors. In the high-power range such a power gain is generally acceptable. Stable, high-gain, low-time-constant preamplifiers of various types are available. The efficiency of the preamplifier will have little effect on the over-all efficiency of the system.

The discussers are correct in qualifying the statement that the motor losses at standstill are zero only when zero torque is produced. Mr. Hollitscher asks for a comparison of the losses with one-phase and with 2-phase control. At maximum torque the losses will be identical for the two systems. But with 10% torque at standstill the loss in the one-phase control will be about 50% of that at 100% while with 2-phase control the loss will be only about 10% of the 100% torque value. Many speed and position control devices operate at relatively low average torques. With 2-phase control the average loss near standstill is proportional to the average torque produced. A decision as to whether auxiliary cooling is required depends on the statistics of the duty cycle for the particular application. If the mean loss is sufficiently low and the variation is sufficiently rapid in comparison with the thermal time constant of the motor, self-cooling is possible. Partial derating of the motor may be necessary in some cases.

We are not as concerned as is Dr. Halacsy about the problem of harmonics in the output wave form. With the duplex doubler arrangement, the only harmonics present are the 6th, 10th, 14th etc. With a resistive load the harmonic content is nowhere greater than about 15% and with an inductive load it is considerably lower. In this motor-drive application, the harmonic power dissipated in the motor has very little effect on the performance or efficiency. Where necessary, the harmonic content can be reduced further by use of more complex connections of the doublers or by use of tuned filters.

In reply to Dr. Manteuffel's comment about the necessity of filtering the output of the preamplifier driving the control circuit; we have applied the unfiltered output of a full-wave rectifier directly to the control windings of a duplex doubler and have found no detectable difference in output characteristics and waveform from that obtained using a pure d-c source.

Magnetic Output Amplifiers for Digital Control

ANDRESS KERNICK
ASSOCIATE MEMBER AIEE

H. A. PERKINS, JR.
ASSOCIATE MEMBER AIEE

WITH the advent of power-frequency magnetic-amplifier devices for logic systems, a need has developed for power-output magnetic amplifiers to raise the very low output level of logic elements to the moderate power requirements of transducer devices. Logic devices generally replace small control relays in industrial applications, but, instead of being able to control loads of approximately 100 to 1,000 volt-amperes, their output is in the range of 0.1 to 1 volt-ampere. Output amplifiers must supply the power required to operate transducer devices such as pneumatic or hydraulic solenoid valves, large power contactors, small motors, etc. Moreover, the output requirements of magnetic amplifiers so applied are unusual and much more stringent in comparison with more common applications such as servomotor amplifiers or proportional-control-device amplifiers. This difficulty in application is principally the result of current design practice for solenoid-operated devices. These solenoids, intended to operate from a power line directly or through a step-down transformer, impose upon the amplifier definite requirements which are discussed. Because a complete line of transducer devices is available, the best interim solution (and possibly the best long-range solution to the problem posed) is to adapt the amplifier to the load rather than to redesign the transducer to be compatible with the amplifier.

Magnetic output amplifiers are considered for control systems containing logic elements of a type introduced by Ramey.[1] Single-phase "doubler" magnetic-amplifier circuits exemplify principles which are also applicable to amplifier circuits controlling certain d-c loads.

Pertinent Electromagnet Characteristics

Following is a review of pertinent characteristics of electromagnets used in transducer devices. A-c electromagnet design concepts as from Roters[2] are to be accommodated by magnetic amplifiers concurrently with NEMA and ASA Standards.[3,4] Magnetic amplifiers and maximum-burdened control transformers are allowed only 5% voltage regulation (between electromagnet open-gap inrush to closed-gap holding) within the NEMA +10% to −15% limits[3] for voltage at the electromagnet coil and the ASA tolerance of ±10% on supply-voltage variation about nominal.[4] Pertinent electromagnet characteristics are:

1. High excitation of open gap.
2. Low excitation of closed gap.
3. Self-regulated armature velocity.
4. Power factor nearly a constant.
5. Coil-voltage tolerance +10% to −15%

Open-gap inrush current is of short duration and normally three to ten times the steady-state excitation of minimum magnetic gap. Closed-gap excitation is limited by temperature rise. Increase of coil impedance with electromagnet armature attraction produces desirable cancellation of velocity and reduces wear at critical closed gaps over the unshaded armature pole face. Accelerated wear at pole contact surfaces, caused by excessive kinetic energy, disrupts designed gap balance between shaded and unshaded portions of the pole face. Unwanted flux shunting of the shading coils results in "chatter" and possible sealed-in by residual flux. Essentially, constant power factor throughout the working cycle happens coincidentally with optimum conditions for minimum force variation designed into shading poles. Characteristics 1, 2, and 4 are corroborated in Table I which demonstrates, too, the tendency for power-factor angle to decrease in accordance with the size of the electromagnet.

Table I. Typical Transducer Characteristics

Transducer	Open-Gap Excitation, Amperes	Open-Gap Power Factor, Per Cent	Inrush-Holding Ratio	Closed-Gap Excitation, Amperes	Closed-Gap Power Factor, Per Cent
Pilot valve *A*	0.362	50.3	1.43	0.253	48.2
Pilot valve *B*	3.56	38.2	7.19	0.495	30.8
Pilot valve *C*	12.3	33.2	6.47	1.90	25.4
Size 1 contactor	1.50	45.7	5.56	0.270	33.6
Size 2 contactor	4.98	29.8	5.72	0.872	21.4
Size 3 contactor	8.00	16.3	5.75	1.39	20.8

Paper 57-774, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted November 16, 1956; made available for printing April 23, 1957.

ANDRESS KERNICK and H. A. PERKINS, JR., are with Westinghouse Electric Corporation, Pittsburgh, Pa.

Helpful comments and suggestions of R. W. Roberts are gratefully acknowledged by the authors.

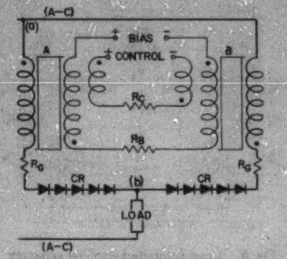


Fig. 1. Conventional doubler magnetic amplifier

Ideal Magnetic-Amplifier Characteristics

1. Flat-voltage regulation over a load-current range of 20 to 1, with the output characteristic lying between 95% to 100% of nominal voltage throughout, is considered sufficient, maximum, obtainable performance.

2. Duration of unregulated output must not exceed a half-cycle at 60 cycles per second because response to changing-load impedance must be such as not to interfere with electromagnet self-regulated armature velocity (characteristic 3).

3. Response at command of control input should be no more than 6 cycles at 60 cycles per second either "on" or "off." "On" and "off" are to be judged by the operated state of a particular transducer. Inherent response of the model transducer "on" is 2 cycles, and "off" is 1 cycle at 60 cycles per second.

4. No tendency to magnify the ±10% variation of line voltage is permitted. An ideal condition is approached if closed-gap excitation voltage remains constant during excursions of line voltage.

5. Flexibility conducive to diversification is desirable. Regulation of output voltage should be accompanied by freedom in the mode of control, e.g., permit design of logic-function devices of the type presented by Radus, Roberts, Rieley, and Fields.[5]

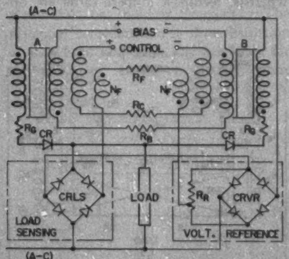


Fig. 2. Negative voltage feedback regulation

Discussion

Review of Present Techniques

Following is a discussion, within the purview of the foregoing requirements, of a-c output circuitry considered.

The simple saturable reactor circuit[6] may be dismissed quickly because examination of its operating characteristics shows that it tends to approximate a current source rather than the voltage source implied in the foregoing. Also, its inherently low gain in comparison with self-saturating circuits would require exhorbitant control power to obtain the desired response time.

Fig. 1 shows the doubler magnetic amplifier: a self-saturating circuit. As shown previously, the voltage regulation of this amplifier with changing load must be no greater than 5% from inrush current to holding current. An excessively large amplifier for the holding-current requirement results. Where a digital load is to be operated, the amplifier must be driven well into saturation by control current so that maximum inrush-current capacity may be realized. A slow speed of response is to be expected in reducing doubler magnetic-amplifier output to the "off" condition. Lord[7] has shown that resetting of a reactor by bias ampere-turns is hindered by a low saturated-impedance drop through the gate winding of the opposite reactor. With the amplifier driven to saturation, reactor resetting is retarded by low initial drop, from point (*a*) to point (*b*) of Fig. 1, comprising the saturated-impedance drop of the opposite reactor and its rectifier forward drop. Not to be forgotten is that resetting of the reactors normally occurs after load current has dropped from the inrush to the holding value, at which current the impedance drops just mentioned are a minimum. Inclusion of several rectifier cells in each gating branch is beneficial. Nonlinear impedance so introduced tends to facilitate the reset operation without greatly increasing output regulation. For this reason use of single-cell silicon diodes as self-saturating rectifiers is particularly undesirable.

Most of the problems set forth in the preceding paragraph may be eliminated or solved in some measure by the introduction of feedback from output to input of the amplifier. Feedback may take the form of positive current feedback or negative voltage feedback.

Fig. 2 shows the circuit of Fig. 1 arranged to have negative voltage feedback. Supply voltage to the gate winding is deliberately made higher for this scheme than for the conventional doubler so that inrush current causes the output to regulate just to the voltage necessary for actuation of the load. After the load has assumed the operated condition, increased impedance of the load will tend to cause load voltage to rise toward the amplifier gate supply voltage. The voltage-feedback circuit consists of feedback windings N_F, resistance R_F, the load-sensing element, and a voltage reference. Ampere-turns are applied which reduce output voltage across the load as a function of the difference between line voltage and amplifier output voltage. *CRLS* provides that feedback be operative only if output voltage is excessive. Obviously, some sophistication of the reference circuit might be made to provide essentially a constant-voltage reference over large limits of changing gate supply voltage instead of reference to voltage that changes directly with line voltage. Clearly any such scheme must effect control of the amplifier most quickly, because self-regulating action, characteristic of a-c electromagnets, may

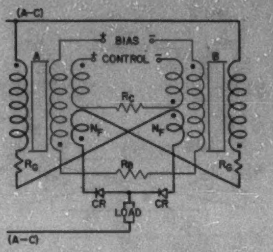


Fig. 3. Positive current feedback regulation

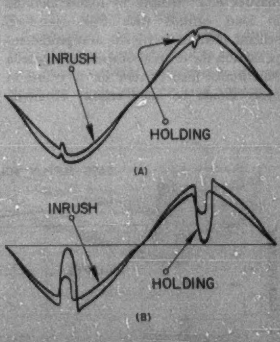


Fig. 4. A—Voltage waveforms without feedback. B—Regulated-voltage waveforms with feedback

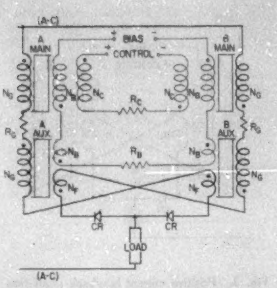


Fig. 5. Load-controlled magnetic amplifier

be largely negated by failure to hold nearly constant voltage throughout the actuation. Since response times for the loads under consideration are typically on the order of a few cycles of the supply frequency, a half-cycle response to feedback control is not only reasonable but probably mandatory. Close control of this sort invites instability problems due to gain and time constants involved in this closed-loop system.

A somewhat simpler scheme as shown in Fig. 3 utilizes positive current feedback. Feedback windings N_F are in series with gate windings of the opposite reactors so that increasing load current causes increasing control ampere-turns to be applied such that output voltage tends to rise. Proper design of the amplifier (regulation characteristics and feedback) should provide that output voltage increase with feedback at the same rate that output voltage would tend to drop due to regulation over the range of load currents required by the electromagnet coil. Ability to follow changes in load current with this open-loop scheme is inherently good, and good compensation for drop in output voltage as a function of load current on a transient basis should be more readily obtained.

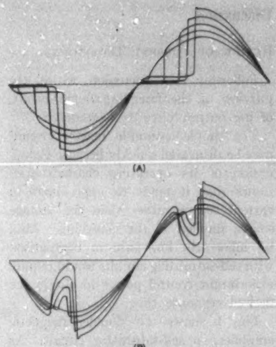


Fig. 6. Voltage waveform family

A—At unity power factor
B—At lagging power factor

In both arrangements (Figs. 2 and 3) improved response may be expected to the input control function over that obtainable with the simple circuit of Fig. 1. Figs. 4(A) and 4(B) make this point clear. Fig. 4(A) compares output waveforms on inrush and holding conditions for the circuit of Fig. 1. Fig. 4(B) compares the same conditions for feedback amplifiers of Figs. 2 and 3. A much higher average voltage drop appears in the gate windings of feedback amplifiers on holding current so that resetting of the reactors by bias ampere-turns is facilitated. High-conductance silicon diodes may be used for *CR*.

Response time for actuating a load to the "on" condition probably differs little from one scheme considered to another. The voltage-feedback system is most likely the fastest because its voltage tends to be excessive during the actuation period. The current-feedback system should be slightly slower because load current must be established before full output can be realized. Two prominent difficulties with these circuits are:

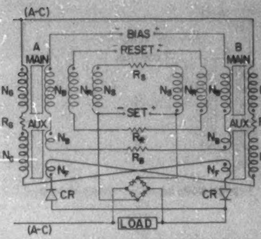


Fig. 8. Load-controlled amplifier MEMORY

1. The entire slope of the input-output transfer characteristic is available for regulating output voltage to a constant value. This is not necessary because, even for a maximum power transfer to the load, amplifier gate-winding drop should not exceed about half the available supply voltage. Thus some means to limit the region of control would be desirable.
2. Summing ampere-turns applied to control the amplifier show that the magnitude of control signal affects the regulation characteristic directly in the circuit of Fig. 3 and to some small extent at least in the circuit of Fig. 2. Considered, therefore, must be any expected variations in control ampere-turns and the critical nature of bias and control settings.

Ideal magnetic-amplifier characteristics compatible with existing transducers are found in the following circuit that overcomes the difficulties just discussed.

Novel Circuit for Digital Loads

Fig. 5 discloses a novel circuit having four saturable magnetic cores for the actuation of digital loads as discussed in the foregoing. Gate windings N_G of main and auxiliary reactors *A* (similarly reactors *B*) are series-connected and so designed that the sum of two voltage drops is approximately equal to the gate supply voltage with the load in the "off" condition. Control ampere-turns are applied only to the main reactors; whereas, feedback ampere-turns are ap-

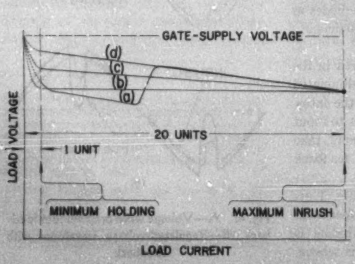
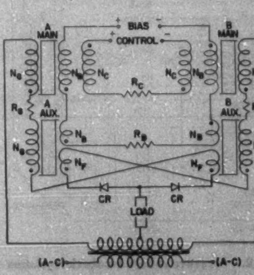


Fig. 7 (left). Voltage-regulation curves at the intrinsic power factor of the saturable reactors

Fig. 9 (right). Load-controlled center tap full-wave magnetic amplifier



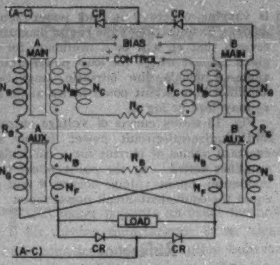


Fig. 10. Load-controlled full-wave bridge magnetic amplifier

plied only to the auxiliary reactors. Sufficient control ampere-turns are provided to drive the main reactors to saturation when the load is to be actuated. Feedback turns on the auxiliary reactors drive them to saturation on maximum inrush-load current, but auxiliary-reactor bias ampere-turns regain control on holding-load current.

Segregation of control and feedback functions makes the regulation characteristic substantially independent of control ampere-turn magnitude. Apportioning gate supply voltage across the reactors in a definite ratio permits the compensation characteristic to be matched to the regulation of the complete amplifier over the contemplated range of load currents.

Fig. 6(A) shows the variation in waveform as the load current is varied from inrush to holding with unity-power-factor load. Fig. 6(B) shows similar waveforms with lagging-power-factor load. Nearly constant load voltage in the "on" condition is obtained with less magnetic material at a given output than for the circuit of Fig. 1.

DESIGN CRITERIA

As shown in Figs. 6(A) and 6(B), the waveform of applied voltage is no longer sinusoidal on holding currents. The principal consideration is that distortion of waveform not cause any tendency for electromagnet armature chatter while holding the load in the "on" condition. Obviously this is largely a function of electromagnet mechanical design. Empirically, a reduction in average output of as much as 25% has not proved excessive. One might arbitrarily provide that a maximum of 25% of the gate supply voltage be dropped across the auxiliary-reactor gate winding.

Having determined this ratio of voltages across the gate windings, the total saturated-impedance drop of both main and auxiliary gate-winding circuits in series may have a maximum of about 25% of the gate supply voltage at inrush-load current. Amplifier size is determined by transducer inrush current at rated operating voltage, provided this voltage is no less than 75% of the gate supply voltage.

Compensation action is a function of the difference between feedback and bias ampere-turns and of linear-slope gain in the auxiliary reactors. Depending upon the particular design, the regulation chactacteristic may follow generally any one of the following three shown in Fig. 7:

1. Overcompensated, curve (*a*).
2. Optimum, curve (*b*).
3. Undercompensated, curve (*c*).

The appendix describes a vector diagram which predicts to a good approximation the conventional doubler magnetic-amplifier voltage regulation with essentially sine waveforms at various power factors. One special case of vector diagram results in the straight-line curve (*d*) of Fig. 7. There exists this one load power factor at which the intrinsic power factor of saturated magnetic-amplifier cores coincides: at this power factor maximum feedback turns are required for curve (*d*) of Fig. 7. At other load power factors regulation is less and, consequently, fewer feedback turns should be used.

Variations of line voltage (limited to ±10%) affect amplifier control, bias, and output equally because it is assumed that all derive their supply from the same source. In the holding-load state, positive current feedback to the auxiliary reactors is of smaller magnitude than bias; therefore, a like per-cent change in bias represents a larger control change than that which occurs by feedback. An inherent ideal regulation exists tending to maintain constant holding-load voltage in spite of line-voltage variation. Auxiliary reactors, operating on the linear slope of their input-output transfer characteristic, have the ability to aid holding at reduced line voltage and to maintain the output voltage by absorbing a greater volt-second area at increased line voltage.

Segregation of main-reactor control from the open-loop feedback to the auxiliary reactors makes performance of logic functions (AND, NOT, OR, and MEMORY)[8]

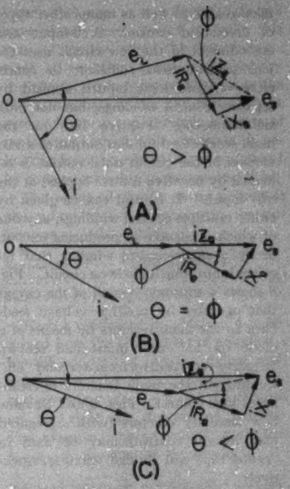


Fig. 12. Vector diagrams having equivalent-circuit power-factor angle θ, related to intrinsic amplifier power-factor angle ϕ, as shown in (A), (B), and (C)

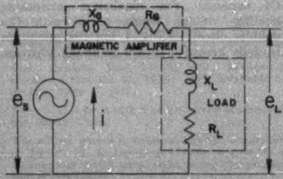


Fig. 11 (left). Thévenin equivalent circuit

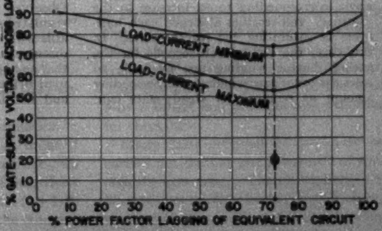


Fig. 13 (right). Load voltage versus equivalent-circuit power factor at constant current manageable as well as many other types of diversified control. A 2-input AND embodiment of the new circuit uses the bias ampere-turns sufficient to cancel either of two logic inputs, applied by separate control windings, but not both simultaneously. Positive bias on the main reactors establishes output of a NOT version of the circuit until output is removed by negative control applied at the NOT input. OR control can be given by either multiple-control windings, any one of which is capable of producing output, or by a single-control winding that receives input from a diode OR circuit. Fig. 8 shows a MEMORY version of the circuit that operates by positive voltage feedback to the main reactors by means of a diode OR for "on" input and seal-in. Negative reset control from a second "off" winding on the main reactors re-establishes bias control. Figs. 9 and 10 show how auxiliary reactors (with segregated main-reactor control) may be used in "center tap" and "bridge" circuits respectively.

Summary

Voltage regulation, response, and control problems are reviewed for magnetic amplifiers operating loads whose range in current requirements may be as great as twenty to one with applicable specifications narrowing output-voltage variation to an allowable 0 to -5%. Conventional feedback circuits are prone to instability and complexity when achieving the following ideal magnetic-amplifier characteristics:

1. Flat (0 to -5%) voltage regulation.
2. Half-cycle response to load changes.
3. Six-cycle response to control.
4. Line-voltage drift not magnified in holding-load current.
5. Diversified noncritical control.

A novel 4-reactor magnetic amplifier is discussed which possesses all of these ideal characteristics.

Appendix

A vector diagram may be constructed which predicts the regulation characteristic of a doubler magnetic amplifier based upon the following assumptions:

1. Sinusoidal voltage and current waveforms.
2. Constant saturated impedance in the magnetic-amplifier gate circuit.

The Thévenin equivalent circuit is shown in Fig. 11. Vector diagrams of the circuit are shown in Fig. 12 for e_s constant and a constant-current load whose power factor is varying, where θ is load power-factor angle. Examination of the vector diagrams reveals that e_L will be a minimum for a given current where the load power-factor angle θ equals the intrinsic magnetic-amplifier gate-circuit power-factor angle ϕ. e_L may lead or lag e_s.

Fig. 13 shows curves of voltage output and equivalent-circuit power factor at constant values of current for a particular amplifier tested. At the point where the maximum dip in voltage occurs, θ and φ equal the equivalent-circuit power factor.

References

1. THE SINGLE-CORE MAGNETIC AMPLIFIER AS A COMPUTER ELEMENT, R. A. Ramey. *AIEE Transactions*, vol. 71, 1952 (Jan. 1953 section), pp. 442–46.

2. ELECTRO-MAGNETIC DEVICES (book), H. C. Roters. John Wiley & Sons, Inc., New York, N. Y., 1941, pp. 446–80.

3. NEMA STANDARDS FOR INDUSTRIAL CONTROL. *IC 1-1954, IC1-11.25B*, National Electrical Manufacturers Association, New York, N. Y., 1954.

4. PREFERRED VOLTAGE RATINGS FOR A-C SYSTEMS AND EQUIPMENT. *ASA C84.1-1954*, American Standards Association, New York, N. Y., 1954, App. A.

5. MAGNETIC AMPLIFIERS FOR REPLACEMENT OF RELAYS, R. J. Radus, R. W. Roberts, O. R. Rieley, C. V. Fields. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 12, 1956, pp. 454–68

6. MAGNETIC AMPLIFIERS (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 77–79, 81.

7. THE INFLUENCE OF MAGNETIC AMPLIFIER CIRCUITRY UPON THE OPERATING HYSTERESIS LOOPS, Harold W. Lord. *AIEE Transactions*, vol. 72, 1953 (Jan. 1954 section), pp. 721–28.

◆

Discussion

H. L. Kellogg (General Electric Company, Fort Wayne, Ind.): The authors have thoroughly reviewed the special magnetic-amplifier characteristics required by the subject application, and have described an effective circuit (Fig. 5) for achieving acceptably flat-voltage regulation. The presentation could perhaps benefit by a more complete explanation of the response-time characteristics of the various circuits.

For example, the feedback circuits shown in Figs. 2 and 3 are described as providing improved response because of the higher voltage drop which appears in the gate windings under holding- or light-load conditions. Presumably, this increased voltage drop results from a delayed firing angle, corresponding to an increase in the "setting" volt-seconds absorbed each half-cycle in saturating the reactor cores. In this case, the setting volt-seconds for the gating reactor winding are available to reset the other gate winding a corresponding amount as required for steady-state conditions. Any additional resetting is ordinarily limited to that corresponding to the forward rectifier drop and the *IR* drops of the gating rectifier and reactor windings. In view of this, it seems doubtful that an appreciable improvement in turn-off response would be obtained.

Equivalent effects would be obtained from the circuit shown in Fig. 5, since the setting volt-seconds of one auxiliary-reactor winding during its gating half-cycle are required to reset the other auxiliary reactor a corresponding amount during steady-state conditions.

In view of the foregoing, a more complete explanation by the authors would be greatly appreciated.

Andress Kernick and **H. A. Perkins, Jr.**: Observed turn-off response in the feedback regulated circuits is superior to that obtainable with the doubler circuit of Fig. 1. The discussion by Mr. Kellogg correctly shows that any improved response must be attributed to increased saturated-impedance drops and not to volt-seconds absorbed by the opposite branch in the prefiring interval. Practically, the saturated impedance for the feedback amplifiers, as a class, would be much higher (approximately five times as shown in the text of the paper). This condition is obtained because the thermal rise on holding current principally determines the physical size of the regulated amplifier along with waveform considerations, whereas only the allowable voltage regulation determines the physical size of the unregulated amplifier.

Another observation might be made to explain the superior response obtainable for the loads discussed. The drop-out voltage is relatively close to the operating voltage of the load. Given a decaying output from the amplifier as a function of time without consideration to any special circuitry, one would expect the load to drop out earlier if the initial voltage existing before the decay is the minimum acceptable holding voltage rather than the maximum. Of necessity the maximum acceptable holding voltage would occur with the doubler circuit of Fig. 1, but the minimum is easily provided by the feedback schemes.

Solution of Simultaneous Equations by Block Elimination

JOHN L. C. LÖF
MEMBER AIEE

THIS paper presents an exact method for solving simultaneous linear algebraic equations which the author believes has many advantages over other procedures. It extends the familiar method of eliminating a single unknown at a time to that of eliminating large groups in a single process. A close relationship to the reduction or "partitioning" of matrices[1,2] is also seen to exist. The process is applicable to small sets of equations to be solved by hand or keyboard as well as to larger sets to be solved by digital computers. The method provides an engineer who needs a quick hand solution to a set, for example, of eight equations in eight unknowns, with a simple procedure which involves only elementary algebra and no complicated rules. There are other methods requiring about the same total amount of work[3-6] and some which break up, as does this method, a large set of equations into several smaller sets,[7,8] but it is felt that the procedure herein presented is more easily learned and more readily applied by a practical engineer. The method is also presented in matrix form in the Appendix.

Method of Solution

In simple elimination methods for solving a set of n simultaneous linear algebraic equations in n unknowns, one of the equations is used to eliminate one unknown from each of the other n-1 equations. This reduces the original set to n-1 equations in n-1 unknowns. The procedure in the block-elimination method is to eliminate as many unknowns at one time as one wishes, i.e., k unknowns may be eliminated so as to reduce the set to the remaining n-k unknowns. Clearly, a considerable amount of work is involved in eliminating a large set of unknowns at one time, but it is fully justified by the resulting simplification.

Consider the following set of n equations in n unknowns:

$$\begin{array}{l}
a_{11}x_1+a_{12}x_2+\ldots+a_{1k}x_k+\ldots+a_{1n}x_n=A_1\\
a_{21}x_1+a_{22}x_2+\ldots+a_{2k}x_k+\ldots+a_{2n}x_n=A_2\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
a_{k1}x_1+a_{k2}x_2+\ldots+a_{kk}x_k+\ldots+a_{kn}x_n=A_k\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
a_{n1}x_1+a_{n2}x_2+\ldots+a_{nk}x_k+\ldots+a_{nn}x_n=A_n
\end{array} \tag{1}$$

All unknowns x_1 through x_k can be eliminated simultaneously by the solution of n-k subsets of k simultaneous equations. These n-k subsets are those obtained directly from equation 1 by rewriting them as

$$\begin{array}{l}
a_{11}y_1+a_{21}y_2+\ldots+a_{k1}y_k=-a_{k+1,1}\;=-a_{k+2,1}=-a_{n1}\\
a_{12}y_1+a_{22}y_2+\ldots+a_{k2}y_k=-a_{k+1,2}\;=-a_{k+2,2}=-a_{n2}\\
\qquad\qquad\qquad\qquad\qquad\text{or}\qquad\qquad\text{etc.}\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
a_{1k}y_1+a_{2k}y_2+\ldots+a_{kk}y_k=-a_{k+1,k}=-a_{k+2,k}=-a_{nk}
\end{array} \tag{2}$$

The notation indicates that each subset is to be solved for y_1 through y_k; first, where the left-hand side of the equations equals the first right-hand column; second, where it equals the second right-hand column; etc., up to the (n-k)th column. Note that the coefficients of the first *row* of equation 2 are the first k coefficients of the first *column* of equation 1, and the coefficients of the second row of equation 2 are those of the second column of equation 1 etc.; i.e., the matrix of equation 2 is the transpose of the matrix of equation 1.

These n-k equations now are solved for the y's. Let the n-k solutions for y_1 through y_k be defined as

$$\begin{array}{l}
y_1=\alpha_1,\beta_1,\gamma_1,\ldots\\
y_2=\alpha_2,\beta_2,\gamma_2,\ldots\\
\cdots\cdots\cdots\cdots\\
y_k=\alpha_k,\beta_k,\gamma_k,\ldots
\end{array} \tag{3}$$

where the α's arise from the a_{k+1} coefficients, the β's arise from the a_{k+2} coefficients, etc.

When all terms of the first equation in equation 1 are multiplied by α_1, those of the second are multiplied by α_2, etc., up to the kth equation by α_k, and then these k equations are added to the (k+1)th equation, the sum of each of the columns 1 through k must be zero since equations 2 have already been solved to make this true. Likewise, when the first k equations of equation 1 are multiplied by the several β's in order and added to the (k+2)th equation, the sum will again be zero, and so on up through n-k sets of multipliers, which will eliminate k unknowns from the original set of equation 1 by successive multiplication and addition.

The result of the foregoing process is the subset of n-k equations in the unknowns x_{k+1} through x_n

$$\begin{array}{l}
b_{k+1,k+1}x_{k+1}+\ldots+b_{k+1,n}x_n=B_{k+1}\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
b_{n,k+1}x_{k+1}+\ldots+b_{n,n}x_n=B_n
\end{array} \tag{4}$$

Each b coefficient in equation 4 is obtained from the original set of equation 1 by the addition of (k+1) original coefficients which have been multiplied by the corresponding α's or β's, and the B terms on the right are determined likewise. For example, the upper-left coefficient $b_{k+1,k+1}$ is the sum of the products

$$\alpha_1a_{1,k+1}+\alpha_2a_{2,k+1}+\ldots+\alpha_ka_{k,k+1}+a_{k+1,k+1}$$

This subset, equation 4 is now solved for x_{k+1} through x_n, possibly by further application of the block-elimination method. Once these unknowns are determined, they may be substituted back into n-k of the original equations to give a final subset of k equations to be solved for the remaining x_1 through x_k.

Simplification and Checking

There are many short cuts used during the foregoing process that are not apparent at first glance. For example, if the subsets of equation 2 are to be solved by block elimination, the set of multipliers to eliminate blocks from equation 2 are identical for all sets of the right-hand terms. Also, when the solutions of equation 4 are substituted back into equation 1, the subset in the remaining variables x_1 through x_k has coefficients identical with those of equation 2 except that they are transposed across the principal diagonal. This simplifies the solution of the final k-by-k subset and provides opportunities for checking.

Many other intermediate checks can be carried out so that one is assured of success the first time. The multipliers, α's, β's, etc., are easily checked by obtaining the sum of the products of the multipliers

---

Paper 57-741, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted October 22, 1956; made available for printing May 17, 1957.

JOHN L. C. LÖF is with The University of Connecticut, Storrs, Conn.

The author wishes to acknowledge the suggestions of Dr. C. G. A. Nordling of The University of Connecticut which were made during the preparation of this paper.

and their corresponding coefficients in rows 1 through k. These sums should, of course, check equation 2 so that only the n-k subset remains. Once the multipliers are correct equation 4 may be checked, if desired, by noting, for example that the sum $b_{k+1,k+1}+\ldots+b_{n,k+1}$ equals the sum of products $a_{1,k+1}(\alpha_1+\beta_1+\ldots)+\ldots+a_{k,k+1}(\alpha_k+\beta_k+\ldots)+a_{k,k+1}+\ldots+a_{n,k+1}$. Since the sums in parantheses are common to each column, this can be performed rapidly for each column. The back substitution of x_{k+1} through x_n into equation 1 is easily checked by comparing the sum of the new right-hand terms with the difference between the sum of the old ones and the sum of all individual x's, each multiplied by the sum of its own coefficients.

Example of General Method

As an example of the general block-elimination method, an 8-by-8 set of nonsymmetrical equations is solved in the following. This solution was obtained with the use of a desk calculator in about 6 hours by an inexperienced operator. For convenience, the equations are given in abbreviated form as follows (equation 5):

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	$= A$	
2.146	3.872	1.964	−0.721	9.830	0.461	−5.837	6.210	28.595	
9.875	8.034	2.169	0.824	−2.631	4.428	3.004	−1.210	9.946	
4.201	2.973	6.241	1.034	0.776	0.467	−3.960	0.388	−8.215	
−3.997	1.236	0.241	4.730	2.604	1.322	3.420	−1.029	46.148	(5)
0.608	4.613	−3.802	5.103	6.078	2.310	1.384	−0.660	72.638	
1.271	2.402	0.607	3.511	2.842	5.211	3.042	1.207	41.342	
0.691	1.370	0.722	0.398	1.207	3.241	3.800	2.614	18.908	
1.204	−0.808	0.261	1.392	2.763	1.313	2.410	4.723	29.840	

This set is reduced by eliminating a block of the first four unknowns x_1 through x_4. This is accomplished by arranging the coefficients of these four unknowns into the array given by equation 2, so that the four numerical subsets in equation 6 are to be solved.

y_1	y_2	y_3	y_4	$=$	or	or	or	
2.146	9.875	4.201	−3.997	−0.608	−1.271	−0.691	−1.204	
3.872	8.034	2.973	1.236	−4.613	−2.402	−1.370	+0.808	(6)
1.964	2.169	6.241	0.241	+3.802	−0.607	−0.722	−0.261	
−0.721	0.824	1.034	4.730	−5.103	−3.511	−0.398	−1.392	

Since the four subsets of equation 6 are still rather large, they are solved by block-elimination of y_1 and y_2, which results in the two subsets of equation 7.

z_1	z_2	$=$	or	
2.146	3.872	−1.964	+0.721	(7)
9.875	8.034	−2.169	−0.824	

It is noted that the coefficients in this array appear in exactly the same locations as they do in the original equation 5, so that the process is actually started by directly copying the foregoing 4-by-2 block, with sign reversal of the latter half, and solving these sets.

The two separate subsets of equation 7 are now solved by whatever method is desired, e.g., by simple elimination or by Cramer's rule. For sufficient accuracy throughout, a few extra figures are carried to give the results of equation 8.

$$\begin{array}{ll} \text{For } z_1: & \lambda_1=+0.351531,\ \mu_1=-0.427865\\ \text{For } z_2: & \lambda_2=-0.702062,\ \mu_2=+0.423347\end{array}\qquad(8)$$

The λ's are the solutions for the z's using the first set of right-hand terms of equation 7, and the μ's are for the second set. The λ's and μ's are the multipliers of equation 6 which eliminate y_1 and y_2 from all four of the subsets. Thus, when the first equation of equation 6 is multiplied by λ_1 and the second by λ_2, and then these two are added to the third, the result is the first equation of equation 9.

Also, when the first equation of equation 6 is multiplied by μ_1 and the second is multiplied by μ_2, and these are then added to the fourth, the second equation of equation 9 results. Note that the solution of only two sets of 2-by-2 equations has resulted in the elimination of two variables from four sets of 4-by-4 each. Equation 9 contains four simple 2-by-2 subsets involving only y_3 and y_4, and furthermore the left-hand sides of the four sets are identical. The four are then readily solved for four sets of y_3 and y_4 to give the solutions of equation 10.

Back substitution of these solutions into the first two equations of equation 6 gives the four more 2-by-2 subsets of equation 11 given below.

$$\begin{array}{cccccccc} y_1 & y_2 & y_3 & y_4 & & \text{or} & \text{or} & \text{or}\\ 0 & 0 & +5.63055 & -2.03182 & =+6.82688 & +0.63256 & -0.00308 & -1.25151\\ 0 & 0 & +0.49515 & +6.96343 & =-6.79576 & -3.98406 & -0.68233 & -0.53479\end{array}\qquad(9)$$

$$\begin{array}{ll} \text{For } y_3: & \alpha_3=+0.838781,\ \beta_3=-0.091762,\ \gamma_3=-0.035008,\ \delta_3=-0.243731\\ \text{For } y_4: & \alpha_4=-1.035565,\ \beta_4=-0.565615,\ \gamma_4=-0.095498,\ \delta_4=-0.059469\end{array}\qquad(10)$$

$$\begin{array}{llll} 2.146y_1+9.875y_2=-8.27087 & -3.14627 & -0.92564 & -0.41778\\ \text{or} & \text{or} & \text{or} & \\ 3.872y_1+8.034y_2=-5.82674 & -1.43009 & +1.14789 & +1.60612\end{array}\qquad(11)$$

It should be observed here that each of the eight right-hand terms of equation 11 is the sum of two products and a third term. Each can, therefore, be obtained by a single continuous operation on a desk calculator by accumulation of products, and no intermediate results are written down. When these four subsets, again with identical left-hand portions, are solved, the eight more solutions for y's are as follows.

$$\begin{array}{ll} \text{For } y_1: & \alpha_1=+0.424345,\ \beta_1=+0.531315,\ \gamma_1=-0.185701,\ \delta_1=+0.915305\\ \text{For } y_2: & \alpha_2=-0.929774,\ \beta_2=-0.434073,\ \gamma_2=-0.053380,\ \delta_2=-0.241218\end{array}\qquad(12)$$

Equations 10 and 12 supply the multipliers that eliminate the four unknowns x_1 through x_4 from the original set of equation 5. Following the foregoing procedure of multiplying the first equation of equation 5 by α_1, etc., up to the fourth by α_4, and adding all of these to the fifth, the first equation of equation 13 is obtained.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	$= B$	
0	0	0	0	10.64983	−2.58872	−10.74915	4.49125	20.84400	
0	0	0	0	7.66280	2.74327	−2.93427	5.57811	26.80934	(13)
0	0	0	0	−0.75384	2.77643	4.53561	1.61007	8.94702	
0	0	0	0	12.05110	0.47440	−2.89546	10.66554	52.87185	

Repetition of the same multiplying process with the β's and addition of all to the sixth equation of equation 5 gives the second equation of 13, etc. The coefficients of the first four unknowns x_1 through x_4 are all listed as zero. Actually,

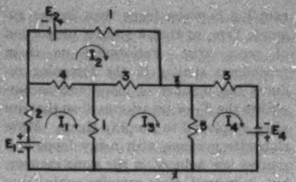


Fig. 1. Four-loop network

with 6-place figures carried in the computation, the largest of these 16 coefficients is 0.00001, and the average of their magnitudes is 0.000003. Thus, these four unknowns have been eliminated successfully with an assurance of negligible error.

Again it is noted that the desk calculator produces each of the terms of equation 13 directly from equations 5, 10, and 12 with no additional recording, since each is an accumulation of four products and a fifth term.

This subset, equation 13, is solved by the block-elimination process in the same way as in the foregoing except that here the amount of work is reduced because there is only one right-hand column in place of the four in equation 6. The solutions for equation 13 are

$$x_5=4.3187, \quad x_7=2.9628,$$
$$x_6=-0.9817, \quad x_8=0.9255 \tag{14}$$

The final step is to back substitute these four values into the original equation 5 to give the sixth 4-by-4 subset in x_1 through x_4 of equation 15.

$$\begin{array}{rrrrr} x_1 & x_2 & x_3 & x_4 & = \\ 2.146 & 3.872 & 1.964 & -0.721 & -1.8592 \\ 9.875 & 8.034 & 2.160 & 0.824 & 17.8753 \\ 4.201 & 2.973 & 6.241 & 1.034 & 0.2660 \\ -3.997 & 1.236 & 0.241 & 4.730 & 27.0198 \end{array} \tag{15}$$

Again, each right-hand of equation 15 is produced by the desk calculator in a single continuous operation of accumulating four products and a fifth term.

These equations are solved in identical fashion to the foregoing. The results of this are

$$x_1=0.6936, \quad x_3=-2.0499,$$
$$x_2=1.3040, \quad x_4=6.0623 \tag{16}$$

As a check upon the accuracy of the work, the solutions shown in equations 14 and 16 are substituted into the original equation 5, with the result that the greatest error between the sum of the left-hand terms with these values substituted and the right-hand terms is 0.0005, and the mean deviation is less than 0.0003. This indicates that carrying the six figures that were used throughout this solution was adequate to provide solutions with maximum errors of roughly 0.05%.

Comments on General Solution

Examination of the foregoing procedure discloses that the solution of a set of equations of any size can be resolved into just two basic operations; one is the solution of 2-by-2 sets, the other is the accumulation of products. In addition, the 2-by-2 sets are also solved by accumulation of products, but here division is also involved. However, if these 2-by-2 sets are solved by Cramer's rule, then only one divisor is required per set. Furthermore, for example, the four solutions of equation 7, the eight solutions of equation 11, and six more involved in obtaining the results in equation 16 are all obtained with an identical divisor. As a result, it is most convenient to perform these divisions by multiplying by the reciprocal of the divisors. Using this procedure in solving the 8-by-8 set it is found that a total of only four divisions is required for the entire process. This is of considerable significance not only when using a desk calculator, but also when programming such a process for digital computers, particularly those of modest speed capabilities.

The work involved here is greatly facilitated by preparing scheduling sheets which outline the procedure. However, this is by no means essential since the process is very simple and there is no need for special formulas or reminders of what is going on. Almost all of the work necessary to the solution of this problem is shown here. In the process, 18 sets of 2-by-2 have been solved using only four different divisors. Additional accumulations of products total 98, with 38 of these being for checking. It is necessary to write down or record only 100 numbers in addition to the given 72 coefficients, and an additional 38 numbers are observed for checking purposes only. Checking occurs at six intermediate stages in the process and finally at the end.

The foregoing method can be modified in a great many ways depending upon the properties of the systems being studied and upon individual preferences. In general, initial elimination of one half of the variables, four in this case, is desirable from the standpoint of organizing the work and providing checking at desirable points. However, circumstances may indicate that other breakdowns could be preferable in many situations. Also, it may be that the variables to be eliminated initially may not appear consecutively in the given array of equations. In this case, one may either rearrange the sequence of the equations and the variables to get the desired block, or leave the equations undisturbed and skip columns or rows or reverse their sequence in any consistent manner that is desired. It is found that substantially the same accuracy will be obtained regardless of the sequencing used.

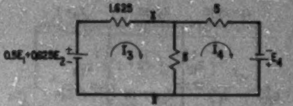


Fig. 2. Equivalent 2-loop network

Special Situations

There are special situations wherein the procedure given here can be shortened considerably. The principal one occurs where several of the coefficients are zero. If only a few are zero, the afore-mentioned general procedure may as well be followed. However, if an appreciable number of coefficients are zero, a substantial saving can be effected. In general, the equations should be rearranged so that as many as possible of the zero coefficients appear in a single block. Then a large number of the a_{k+1}, a_{k+2}, etc., terms on the right-hand side of the n-k subsets of equation 2 can be zero. If $a_{k+1,1}$ through $a_{k+1,k}$ are all zero, the number of k-by-k subsets in equation 2 to be solved is reduced by one since the resulting α's are all zero by inspection. If $a_{k+2,1}$ through $a_{k+2,k}$ are all zero, the β's are all zero, and one more subset need not be solved, etc. Much time may be saved by selecting the original subset of variables to be eliminated so as to take advantage of such situations, especially when solving equations by hand calculation.

Further simplification is achieved in the special cases where the equations can be arranged so that all of the coefficients either above or below the diagonal on the left-hand side of equation 2 are zero. Such situations might arise even though the original equation 1 could not be arranged so that all terms either above or below the diagonal are zero. In these special cases, α, β, etc., are then found by simple direct calculation and back substitution rather than by elimination.

In the case of equations that are symmetrical about the diagonal, the amount of work is again reduced since the result-

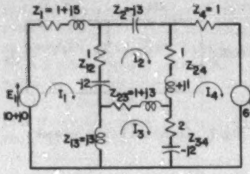


Fig. 3. Four-loop a-c network

ing subset of n-k equations in equation 4 must also be symmetrical. This means that it is only necessary when determining equation 4 to compute the diagonal terms, and either those above the diagonal or those below. In addition, the back-substitution of the variables x_{k+1} through through x_n is simplified since the resulting subset in x_1 through x_k is also symmetrical so that again it is necessary to compute only the diagonal coefficients and either those above or those below. Furthermore, the solution of all of the subsets involved in the foregoing by the block-elimination method is likewise simplified, since all of the subsets are also symmetrical.

Application to Electric Networks

The simple electric network shown in Fig. 1 provides an example of a system symmetrical about the principal diagonal wherein some of the coefficients are zero. The procedure is carried only far enough here to illustrate the reduction of work in such systems. A primary purpose of this section is to illustrate some interesting possibilities for network reduction and analysis by this method.

The 4-loop equations describing the behavior of this d-c network are given in symbolic form by equation 17.

$$\begin{array}{rrrrl} I_1 & I_2 & I_3 & I_4 = & \\ 7 & -4 & -1 & 0 & E_1 \\ -4 & 8 & -3 & 0 & E_2 \\ -1 & -3 & 9 & -5 & 0 \\ 0 & 0 & -5 & 10 & E_4 \end{array} \tag{17}$$

It is noted that the voltage sources are left as symbols rather than given numerical values. Following the procedures foregoing, current unknowns I_1 and I_2 are eliminated in a block by solving the sets of equation 18 which arise from the first two columns of equation 17.

$$\begin{array}{rrrr} \alpha_1 & \alpha_2 = & & \\ 7 & -4 & +1 & 0 \\ & & & \text{and} \\ -4 & 8 & +3 & 0 \end{array} \tag{18}$$

Solution of this set for the α's quickly gives the multipliers to accomplish the elimination:

$\alpha_1 = +0.5,\ \alpha_2 = +0.625$

Of course, there need be no second set of β multipliers in this case where zeros occur at the bottom of the first two columns of equation 17. When the first equation of equation 17 is multiplied by α_1 and the second is multiplied by α_2 and the two equations are then added to the third, the set of equation 19 results.

$$\begin{array}{l} 6.625I_3 - 5I_4 = 0.5E_1 + 0.625E_2 \\ -5I_3 + 10I_4 = E_4 \end{array} \tag{19}$$

This 2-by-2 set is seen to be symmetrical, but this occurs automatically in this case where several zeros exist. However, this situation will occur with all symmetrical sets regardless of the presence or absence of zeros. This can be proved by examining the formation of the new coefficients in the resulting subset. This symmetry then contributes to an appreciable reduction in the number of terms to be computed.

Once the voltage sources are given numerical values, I_3 and I_4 can be determined. However, the remainder of the work is straightforward and need not be done here. Of more interest is the new circuit, represented by the equations in equation 19, which is shown in Fig. 2. This 2-loop system is identical with the original to the right of points x-x. To the left of x-x is the *Thévenin equivalent* of the original circuit, and it is obtained here in a very simple manner. The equivalent voltage source is composed of the linear superposition of the original sources E_1 and E_2, and the proportionality factors are the α's determined previously, i.e.,

$$E_{eq} = \alpha_1 E_1 + \alpha_2 E_2 \tag{20}$$

This procedure thus allows one to retain symbolic representation of these right-hand terms without complicating the work in any way. The process can be carried on further by simple elimination of I_3 to get the single-loop representation of the system with the linear combination of all three voltage sources acting in this loop.

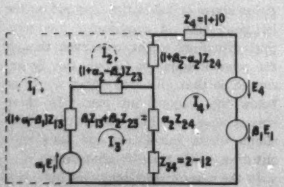

$Z_4 = 1+j0$
I_2
$(1+\beta_2-\alpha_2)Z_{24}$
$(1+\alpha_2-\beta_2)Z_{23}$
I_1
E_4
I_4
$(1+\alpha_1-\beta_1)Z_{13}$
$\beta_1 Z_{13}+\beta_2 Z_{23} = \alpha_2 Z_{24}$
$\beta_1 E_1$
I_3
$\alpha_1 E_1$
$Z_{34} = 2-j2$

Fig. 4. Equivalent 2-loop circuit of Fig. 3

Ladder networks have somewhat looser coupling between loops than in the example here, so that a greater proportion of zeros exist. Performing on them procedures similar to the foregoing reduces the number of loops and accomplishes the $T-\pi$ transformations that are often employed to simplify them. More elaborate systems with many loops and sources are reduced in the same manner, but, in general, more complex equivalents result initially rather than simple Thévenin equivalents. Nevertheless, these equivalents can be exceedingly useful when, for example, a study of a power system is to be made wherein several sources are to be held fixed while the effect of changing a few others is to be determined.

It is of considerable interest to observe what does occur when this process is applied to a general network. Each equation of equation 2 may be interpreted as equating the *voltage drops* in the corresponding loop to a voltage which equals the *mutual impedance* between this loop and an eliminated loop in the system, with its own voltage source equal to zero and the remainder of the circuit open-circuited. (The dimensions of this equation may be resolved by mentally multiplying the mutual impedance by 1 ampere.) The resulting α's, β's, etc., in equation 3 are then the open-circuit currents in the corresponding mutual impedances if voltages equal to these impedances were substituted for the actual voltages. The actual current in each loop under open-circuit conditions equals the product of the foregoing hypothetical current multiplied by the ratio of the actual applied voltage in that loop to the afore-mentioned mutual impedance. The open-circuit voltage across each mutual impedance is simply the product of the hypothetical current α, β, etc., and the corresponding impedance. The sum of the $k+1$ terms giving B_{k+1} of equation 4 then is the equivalent superposition of all of the open-circuit voltages which effect the $k+1$st loop in the new reduced network. The B's therefore give several equivalent sources that are applied to the remaining network.

The b coefficients of the reduced network of equation 4 become the new self- and mutual impedances of this network. Each one represents the total of an original impedance of this block and the sum of proportionate parts of all the impedances from the eliminated block that are mutual to its own loop. It was seen that the α's, β's, etc., indicate the proportionate influence of each source voltage on the open-circuit voltage across any mutual impedance, as for example

Vo
Pt.
19

in equation 19. Likewise, these same α's etc., also indicate the influence of the mutual impedances of the original loops on the new self- and mutual impedances of the reduced system. This results in the new equivalent internal impedances, the b's of equation 4, that are the sums of products of the nondimensional influence coefficients α, β, etc., and the original mutual impedances.

$$\begin{array}{ccccc} I_1 & I_2 & I_3 & I_4 & = \\ 0 & 0 & +(4.56+j0.92) & -(1.56-j1.08) & 5.6-j0.8 \\ 0 & 0 & -(1.56-j1.08) & +(4.16-j1.48) & 6.4+j6.8 \end{array} \qquad (24)$$

Complex Coefficients

The simultaneous equations most commonly encountered by electrical engineers have complex coefficients and are symmetrical about the principal diagonal. These can be solved by the foregoing method with no increase in complexity other than that which ordinarily arises from dealing with the complex numbers themselves.

It is noted again that only a small number of divisions need be performed; this is of increased importance when dealing with complex numbers which imply the rationalization process for division. In general, the rectangular form of expressing complex numbers is found preferable to the polar form if good accuracy is required.

As an example involving complex coefficients, the currents in the 4-loop network of Fig. 3 are determined. The equations for this network are given by equation 21 in symbolic form.

$$\begin{array}{ccccl} I_1 & I_2 & I_3 & I_4 & = \\ 2+j6 & -(1-j2) & -j3 & 0 & E_1=10+j0 \\ -(1-j2) & 3-j1 & -(1+j3) & -(1+j1) & 0 \\ -j3 & -(1+j3) & 3+j4 & -(2-j2) & 0 \\ 0 & -(1+j1) & -(2-j2) & 4-j1 & E_4=6+j8 \end{array} \qquad (21)$$

Following the previous procedures, the first 2-loop currents are eliminated by arranging the coefficients of the first two columns into the equations of equation 22.

$$\begin{array}{ccccc} y_1 & y_2 & = & & \\ 2+j6 & -(1-j2) & +j3 & & 0 \\ & & & \text{or} & \\ -(1-j2) & 3-j1 & +(1+j3) & & +(1+j1) \end{array} \qquad (22)$$

Solving these two sets by Cramer's rule requires only one divisor, and rationalization of this divisor gives in its place a multiplier. By leaving the numbers in the rectangular form the solution of equation 22 then results again in the accumulation of products which is carried out in simple fashion on a desk calculator to give the following.

For y_1: $\alpha_1=0.56-j0.08$,
$\beta_1=0.04-j0.12$

For y_2: $\alpha_2=0.24+j0.68$,
$\beta_2=0.16+j0.32$ (23)

These multipliers eliminate the first two currents by application of the multiplication and addition process to give equation 24.

The solution of equation 24 is given later, but at this point it is again appropriate to interpret further the process involved.

If the coefficients of equation 24 were left in their symbolic form using the original impedances of the network and the foregoing multipliers, the 2-loop circuit of Fig. 4 would represent these equations. This circuit is drawn so as to emphasize the relation between it and the original circuit. It can be seen that the new impedances of this circuit are in accordance with the discussion in the previous section. The new mutual impedance between loops 3 and 4 consists of the original Z_{34} and an additional impedance caused by the mutual impedances from loops 1 and 2 to either loop 3 or loop 4. This additional term can be expressed either as $\beta_1 Z_{13}+\beta_2 Z_{23}$ or as $\alpha_1 Z_{14}+\alpha_2 Z_{24}$; however, the Z_{14} is zero in this case. The new total impedance in each loop is the sum of the original loop impedance and additional parts of those eliminated impedances which are mutual to the loop. For example, the new total impedance in loop 3 is

$$Z_{33}+Z_{13}+Z_{23}+\alpha_2 Z_{23}=4.56+j0.92$$

New voltage sources are inserted into each loop which account for the proportionate effects of the eliminated source in the remaining loops. These are the open-circuit voltages appearing in these loops which result from all the given sources.

The solutions of equation 24 are obtained with only one new complex divisor which is rationalized into a multiplier. These are then back substituted into equation 21 to obtain the equations for the first two currents, and this yields the equations of equation 25.

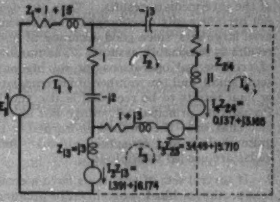


Fig. 5. A second 2-loop equivalent of Fig. 3

$(2+j6)I_1-(1-j2)I_2=$
$11.3914+j6.1738$

$-(1-j2)I_1+(3-j1)I_2=$
$3.5866+j8.8750$ (25)

The solution equation 25 employs the same divisor as does equation 22. Thus, the solution of the four equations of equation 21 is obtained by employing only two rationalization processes. The solutions of equations 24 and 25 are found, with somewhat greater precision than is frequently justified, to be

$$\begin{aligned} I_1&=1.76304-j1.76195 \\ I_2&=0.18904+j1.25867 \\ I_3&=2.05792-j0.46380 \\ I_4&=1.65119+j1.51386 \end{aligned} \qquad (26)$$

Back substitution of these solutions into equation 21 for checking purposes establishes that the maximum error in reproducing the right-hand terms of equation 21 is 0.00003.

Before concluding, the circuit represented by equation 25 as given in Fig. 5 may be examined. Of course, all of the impedances here are identical with the corresponding ones of Fig. 3, but new voltage sources are inserted into the two loops. These sources merely are the effects of neighboring currents, now eliminated, passing through the impedances mutual to the loops of interest. This circuit is rather obvious by inspection, but this is in contrast to the more involved circuit of Fig. 4. On the other hand, if the third- and fourth-loop currents were eliminated first, then the circuit involving the first two currents would have a character similar to that illustrated in Fig. 4.

Conclusions

The method for solving simultaneous equations presented here possesses several advantages which are summarized in the following:

1. It is a rapid, well-organized method in which it is easy to keep track of the progress in the work and also to check the accuracy

of the intermediate results. Thus one does not become easily discouraged when solving even large sets, since good accuracy of final results is reasonably assured. The number of arithmetical operations for solving any set is small, and the number of intermediate or final results to be recorded is also small. Perhaps its principal advantage lies in the fact that once the method is learned it is quite obvious, and no special rules or formulas are needed to apply it.

2. If several sets of equations have a common section or block wherein corresponding coefficients are alike in each set, this method gives a means of eliminating all the unknowns which correspond to this block by determining only a single group of multipliers for all sets. The several reduced sets of equations are then solved separately with far less work than that required for solving each large set from the beginning. This has great importance where a small section of a large electric network is to be changed a number of times and solutions are desired for each.

3. The method takes full advantage of the presence of zeros in any of the coefficients and also in the forcing functions. Additional savings in work occur where a set of equations is symmetrical about the principal diagonal.

4. Complex coefficients can be handled by procedures which are identical with those employed for real coefficients. The rectangular form is preferable to polar in order to obtain accurate results when using any type of machine calculator.

5. The block-elimination process provides a method to be considered for use on digital computers. It has the particular advantage that one good program for solving an n-by-n set will quickly lead to the solution of larger sets by the elimination of n variables at a time. This makes it possible for small computer installations to cope with much larger sets of equations. The storage capacity required by this method is quite small since the number of stored intermediate and final results is little more than the number of given coefficients.

6. The solution of equations describing the behavior of electric networks is a relatively simple process by this method and the likelihood of error is small. Solutions for linear circuits involving either direct or alternating currents and bilateral or unilateral elements can be handled with nearly identical techniques. Perhaps of greater advantage than this is the application to network analysis itself. Large networks are simplified in an orderly fashion to small ones in such a way that the new components are clearly related to the original ones in a predictable manner.

7. The method establishes that one large set of simultaneous equations is equivalent to a certain number of smaller sets. For example, the solution of ten equations in ten unknowns can involve the solution of seven sets of five equations in five unknowns. A set of $(p+q)$ equations in $(p+q)$ unknowns is the equivalent of $(p+1)$ sets of q equations in q unknowns and one set of p equations in p unknowns.

Appendix

To achieve greater compactness of mathematical expression the previously described procedure for solving simultaneous equations can, of course, be expressed in matrix notation by following the simple foregoing pattern. Matrices of interest are defined in the following.

$$A=\begin{Vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{Vmatrix}=\left\|\begin{array}{l|l} a_{11}\ldots a_{1k} & a_{1,k+1}\ldots\ldots a_{1n} \\ \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots\ldots \\ a_{k1}\ldots a_{kk} & a_{k,k+1}\ldots\ldots a_{kn} \\ \hline a_{k+1,1}\cdot a_{k+1,k} & a_{k+1,k+1}\ldots a_{k+1,n} \\ \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots\ldots \\ a_{n1}\ldots a_{nk} & a_{n,k+1}\ldots a_{nn} \end{array}\right\| \tag{27}$$

where A is partitioned into four submatrices. The square submatrix of order k, A_{11} contains the first k coefficients of each of the k unknowns to be eliminated. A_{22} is also square whereas A_{12} and A_{21} may or may not be square.

The multipliers in equation 3 are put in the matrix

$$\alpha=\begin{Vmatrix} \alpha_1\alpha_2\ldots\alpha_k \\ \beta_1\beta_2\ldots\beta_k \\ \gamma_1\gamma_2\ldots\gamma_k \\ \ldots\ldots\ldots \end{Vmatrix} \tag{28}$$

Column matrices for the unknowns are

$$X=\begin{Vmatrix} X_1 \\ X_2 \end{Vmatrix}=\begin{Vmatrix} x_1 \\ \vdots \\ x_k \\ \ldots \\ x_{k-1} \\ \vdots \\ x_n \end{Vmatrix} \text{ and } C=\begin{Vmatrix} C_1 \\ C_2 \end{Vmatrix}=\begin{Vmatrix} c_1 \\ \vdots \\ c_k \\ \ldots \\ c_{k-1} \\ \vdots \\ c_n \end{Vmatrix} \tag{29}$$

where the unknowns x_1 through x_k contained in X_1 are to be eliminated first.

Equation 1 is then expressed in matrix form as

$$A.X=C \tag{30}$$

where the c's here are the equivalent of the knowns noted as A_1 through A_n in equation 1. The matrix A may be partitioned into its submatrices and then equation 30 may be expanded into

$$\begin{Vmatrix} A_{11} & 0 \\ A_{21} & 0 \end{Vmatrix}\cdot\begin{Vmatrix} X_1 \\ X_2 \end{Vmatrix}+\begin{Vmatrix} 0 & A_{12} \\ 0 & A_{22} \end{Vmatrix}\cdot\begin{Vmatrix} X_1 \\ X_2 \end{Vmatrix}=\begin{Vmatrix} C_1 \\ C_2 \end{Vmatrix} \tag{31}$$

where the zeros represent the corresponding submatrices with zeros substituted for the a's. Equation 31 is multiplied throughout by a specially selected square matrix of the same order as A

$$\begin{Vmatrix} 0 & 0 \\ \alpha & U \end{Vmatrix}\cdot\begin{Vmatrix} A_{11} & 0 \\ A_{21} & 0 \end{Vmatrix}\cdot\begin{Vmatrix} X_1 \\ X_2 \end{Vmatrix}+\begin{Vmatrix} 0 & 0 \\ \alpha & U \end{Vmatrix}\cdot\begin{Vmatrix} 0 & A_{12} \\ 0 & A_{22} \end{Vmatrix}\cdot\begin{Vmatrix} X_1 \\ X_2 \end{Vmatrix}=\begin{Vmatrix} 0 & 0 \\ \alpha & U \end{Vmatrix}\cdot\begin{Vmatrix} C_1 \\ C_2 \end{Vmatrix} \tag{32}$$

where α is defined in equation 28 and U is the unit diagonal matrix of order n-k.

$$U=\begin{Vmatrix} 1\ldots0 \\ \ldots\ldots \\ 0\ldots1 \end{Vmatrix} \tag{33}$$

The α's of the selected matrix are defined such that

$$\begin{Vmatrix} 0 & 0 \\ \alpha & U \end{Vmatrix}\cdot\begin{Vmatrix} A_{11} & 0 \\ A_{21} & 0 \end{Vmatrix}=0 \tag{34}$$

or simply

$$\alpha.A_{11}+A_{21}=0 \tag{35}$$

so that

$$\alpha=-A_{21}.A_{11}^{-1} \tag{36}$$

where A_{11}^{-1} is the inverse of A_{11}, i.e., $A_{11}.A_{11}^{-1}=1$.

This definition of α removes the left-hand product from equation 32 and reduces it to

$$\begin{Vmatrix} 0 & 0 \\ \alpha & U \end{Vmatrix}\cdot\begin{Vmatrix} 0 & A_{12} \\ 0 & A_{22} \end{Vmatrix}\cdot\begin{Vmatrix} X_1 \\ X_2 \end{Vmatrix}=\begin{Vmatrix} 0 & 0 \\ \alpha & U \end{Vmatrix}\cdot\begin{Vmatrix} C_1 \\ C_2 \end{Vmatrix} \tag{37}$$

or

$$(\alpha.A_{12}+A_{22}).X_2=\alpha.C_1+C_2 \tag{38}$$

Equation 38 expresses in matrix form the same thing as equation 4 in the form of simultaneous equations. The solution is found by multiplying equation 38 by the inverse of $\alpha.A_{12}+A_{22}$

$$X_2=(\alpha.A_{12}+A_{22})^{-1}.(\alpha.C_1+C_2) \tag{39}$$

Equation 39 is a familiar result obtained from the partitioning of matrices.[1,2] The matrix X_1 may be found by an identical process with one's substituted for two's and vice versa, or by substitution of equation 39 in equation 31. The matrix method gives, therefore, a more formal presentation of the simple algebraic process of block elimination.

References

1. Tensor Analysis of Networks (book), G. Kron. John Wiley & Sons, Inc., New York, N. Y., 1939, pp. 242–69.

2. The Mathematics of Circuit Analysis (book), E. A. Guillemin. John Wiley & Sons, Inc., 1949, pp. 30–71.

3. A Short Method for Evaluating Determinants and Solving Systems of Linear Equations with Real or Complex Coefficients, Prescott D. Crout. *AIEE Transactions*, vol. 60, 1941, pp. 1235–41.

4. Linear Computations (book), Paul S. Dwyer. John Wiley & Sons, Inc., 1951.

5. Solving Linear Algebraic Equations Can Be Interesting, G. E. Forsythe. *Bulletin*, American Mathematical Society, Providence, R. I., vol. 59, 1953, pp. 299–329; bibliography.

6. Contributions to the Solution of Series of Linear Equations and the Determination of Eigenvalues, O. Taussky, editor. *Applied Mathematics, Series 39*, National Bureau of Standards, Washington, D. C., 1954, pp. 1–54.

7. A Method of Solving Very Large Physical Systems in Easy Stages, G. Kron. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 42, 1954, pp. 680–86.

8. Analysis of Rigid-Jointed Plane Frameworks, R. K. Livesley, T. M. Charlton. *Engineering*, London, England, Feb. 1954, pp. 239–41.

Operational Magnetic Amplifiers with Multiple Control Windings

T. T. KADOTA
STUDENT MEMBER AIEE

H. C. BOURNE
MEMBER AIEE

ALTHOUGH a great deal of effort has been expended in the development of analyses of magnetic amplifiers, a completely satisfactory mathematical representation of this dynamic operation, characterized by the soundness of its mathematical basis and by the convenience of its application in analysis and synthesis, has yet to appear. Because of the discrete nature of the operation, the input and output of the magnetic amplifier must be expressed in terms of discrete functions whose variable is the number of half-cycles of the supply voltage frequency. As a result, the dynamic operation must be described by a difference equation which relates input and output.

There are two ways often used to obtain the transfer function of magnetic amplifiers. One is to develop an approximate differential equation which is analogous to the difference equation and to obtain the transfer function by the Laplace transform method. The other is to apply the Laplace transform directly to the difference equation and to obtain the transfer function with the approximation, $e^{Ts}=1+Ts$. The first method is physically improper for approximating the discrete nature of operation with continuous functions, and the second method is mathematically difficult in taking the Laplace transform of discrete functions. However, the degree of approximation is same in both cases.

The objective of the first part of this paper is to eliminate these difficulties and to establish a consistent mathematical analysis without excessive complication. For this purpose, application of the z-transform method through the impulse function representation seems most appropriate.

In control systems it is frequently required to control a single output quantity with multiple input quantities. The control of magnetic amplifiers with multiple inputs can be easily realized by arranging the necessary number of control windings on the cores. In a more general case, some or all of the control circuits include frequency sensitive components. Some of them can be used as feedback circuits.

If frequency sensitive components are represented by their general circuit parameters in operational forms, voltages and currents in the circuit equations and the magnetic constraint equations of the magnetic amplifier must be expressed in operational forms accordingly. To avoid excessive complication in analysis certain restrictions may be imposed upon frequency sensitive components. Yet, these restrictions must be such that the characteristics of the frequency sensitive components may appear in the operating characteristics of a magnetic amplifier without destroying its basic modes of operation.

The second part of this paper is devoted to the derivation of the output equations of the magnetic amplifier first for n-resistive control circuits and then for n-frequency sensitive control circuits.

Application of the z-Transform Method to the Difference Equation of Magnetic Amplifiers

Under certain practical assumptions concerning characteristics of cores and rectifiers,[1] the dynamic operation of typical magnetic amplifiers; such as, series-connected saturable reactors, self-saturating magnetic amplifiers, and the half-cycle response circuits, is described by the following linear first-order difference equation.

$$\bar{e}_L(n+1)=K_1\bar{e}_c(n+1)+K_2\bar{e}_c(n)+K_3\bar{e}_L(n)+K_4 \quad (1)$$

in which $\bar{e}_c(n)$ and $\bar{e}_c(n+1)$ (or $\bar{e}_L(n)$ and $\bar{e}_L(n+1)$) are half-cycle average values of the control (or output) voltages for the nth and the $(n+1)$th half-cycle respectively. The constants K_1, K_2, K_3, and K_4 are functions of the supply voltage, the magnetizing currents (static and dynamic) and the circuit parameters and configuration. For a series-connected saturable reactor circuit K_1, K_2, K_3, and K_4 are all nonzero; for the self-saturating magnetic amplifier circuits K_1 is zero and all the other K's are nonzero; for the half-cycle response circuits K_1 and K_3 are zero and K_2 and K_4 are nonzero.

Equation 1 consists of discrete functions whose variable is the number of half-cycles of the supply voltage frequency. In order to apply the Laplace transform for derivation of the transfer function these discrete functions must be replaced by Laplace transformable functions in a unique manner. One of the ways of replacing them is the impulse function representation. The impulse function representation consists of replacing a discrete function by the corresponding impulse function in the following manner; $\bar{e}(n)\rightarrow e^*(t)$, $\bar{e}(n+1)\rightarrow e^*(t+T)$. The impulse function $e^*(t)$ has an area corresponding to the value of $\bar{e}(n)$; a constant is interpreted as a train of unit-impulses at nT multiplied by the constant $T=\pi/\omega_s$ (the sampling period) and ω_s is the supply voltage frequency.

Application of the impulse function representation to equation 1 results in the following new difference equation in terms of impulse functions.

$$e_L^*(t+T)=K_1e_c^*(t+T)+K_2e_c^*(t)+K_3e_L^*(t)+K_4\delta_T(t) \quad (2)$$

The Laplace transform of equation 2 is

$$e^{Ts}[E_L^*(e^{Ts})-e_L(0)]=K_1e^{Ts}[E_c^*(e^{Ts})-e_c(0)]+K_2E_c^*(e^{Ts})+K_3E_L^*(e^{Ts})+K_4\frac{e^{Ts}}{e^{Ts}-1}$$

Replacing e^{Ts} by z and solving for the transformed output voltage,

$$E_L^*(z)=\frac{K_1z+K_2}{z-K_3}E_c^*(z)+\frac{K_4}{z-K_3}\,\frac{z}{z-1}+\frac{z}{z-K_3}[e_L(0)-K_1e_c(0)] \quad (3)$$

If the magnetic amplifier is assumed to be initially in steady-state with no control voltage applied, the initial condition is obtained by substituting $n+1=0$ into equation 1. Thus,

$$\bar{e}_L(0)=K_1\bar{e}_c(0)+\frac{K_4}{1-K_3} \quad (4)$$

Inserting the initial condition, equation 4, into equation 3 the output equation is obtained as

$$E_L^*(z)=\frac{K_1z+K_2}{z-K_3}E_c^*(z)+\frac{K_4}{1-K_3}\,\frac{z}{z-1} \quad (5)$$

The incremental transfer function is

Paper 57-772, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted October 29, 1956; made available for printing April 24, 1957.

T. T. KADOTA is a student and H. C. BOURNE is on the faculty at the University of California, Berkeley, Calif.

This research was supported by the Air Force Cambridge Research Center, Air Research and Development Command under Contract *AF 19(604)-1813*. The authors wish to thank E. I. Jury and D. Nitzan for their valuable assistance.

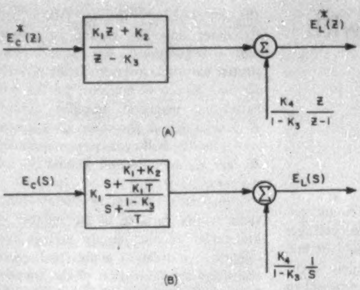

$E_c^*(z)$
$\frac{K_1 z + K_2}{z - K_3}$
Σ
$E_L^*(z)$
$\frac{K_4}{1-K_3}\ \frac{z}{z-1}$
(A)
$E_c(S)$
$K_1 \frac{S+\frac{K_1+K_2}{K_1T}}{S+\frac{1-K_3}{T}}$
Σ
$E_L(S)$
$\frac{K_4}{1-K_3}\ \frac{1}{S}$
(B)

Fig. 1. Block diagrams of equations 5 and 5(A)

A—The magnetic amplifier described by equation 5
B—The approximate continuous representation of the magnetic amplifier described by equation 5(A)

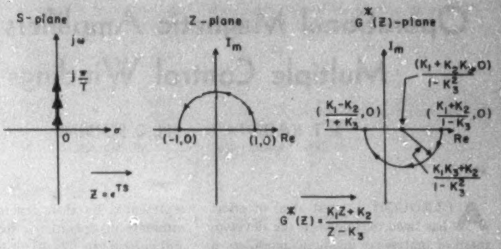

S-plane
jω
$j\frac{\pi}{T}$
0
σ
$z=\epsilon^{Ts}$
Z-plane
Im
(-1,0)
(1,0) Re
$G^*(Z)$-plane
Im
$(\frac{K_1+K_2K_3}{1-K_3^2},0)$
$(\frac{K_1-K_2}{1+K_3},0)$
$(\frac{K_1+K_2}{1-K_3},0)$
Re
$\frac{K_1K_3+K_2}{1-K_3^2}$
$G^*(Z)=\frac{K_1Z+K_2}{Z-K_3}$

Fig. 2. Graphical illustration of the successive mapping on complex planes and the Nyquist plot of G*(z)

given as the following equation

$$G^*(z)=\frac{K_1z+K_2}{z-K_3}=K_1+\frac{K_1K_3+K_2}{z-K_3} \tag{6}$$

The continuous approximation of equations 5 and 6 are obtained by changing $E^*(z)$ to $E(s)$.

$$E_L(s)=K_1\frac{s+\frac{K_1+K_2}{K_1T}}{s+\frac{1-K_3}{T}}E_c(s)+\frac{K_4}{1-K_3}\frac{1}{s} \tag{5A}$$

$$G(s)=K_1\frac{s+\frac{K_1+K_2}{K_1T}}{s+\frac{1-K_3}{T}}=K_1+\frac{\frac{K_1K_3+K_2}{1-K_3}}{1+\frac{T}{1-K_3}s} \tag{6A}$$

The block diagram interpretation of equations 5 and 5(A) is given in Fig. 1.

Application of the z-transform[2] method to the transfer function, equation 6, results immediately in the following conclusions concerning the operating characteristics.

1. The stability criteria are given as

$$|K_3|<1 \tag{7}$$

This result is expected from the difference equation, equation 1, since K_3 is the feedback factor (the coefficient of $e_L(n)$).

2. The steady-state gain is given by

$$K_{vD-C}=\lim_{z\to 1}G^*(z)=\frac{K_1+K_2}{1-K_3} \tag{8}$$

3. The transient response consists of two types of responses, namely, instantaneous and exponential. The time constant of the exponential response is

$$\tau=-\frac{1}{\ln|K_3|}T \tag{9}$$

If $K_1=0$, in the case of the self-saturating magnetic amplifiers, the transient response is purely exponential and the gain-bandwidth product is

$$\frac{K_{vD-C}}{\tau}=-\frac{K_2}{1-K_3}\ \frac{\ln|K_3|}{T} \tag{10}$$

If K_1, $K_3=0$, in the case of the half-cycle response circuits, the time constant of the exponential response becomes zero. However, the half-cycle response should be differentiated from the instantaneous response.

The continuous approximations of the operating characteristics are obtained from equation 6(A) by applying the ordinary Laplace transform method as follows:

1. The stability criteria are given as

$$\frac{1-K_3}{T}>0,\text{ or }K_3<1 \tag{7A}$$

2. The d-c steady state gain is given by

$$K_{vD-C}=\lim_{s\to 0}G(s)=\frac{K_1+K_2}{1-K_3} \tag{8A}$$

3. The transient response consists of an instantaneous and an exponential response, of which the time constant is

$$\tau'=\frac{1}{1-K_3}T \tag{9A}$$

It should be noticed that $1-K_3$ is the first term of the expansion series of $\ln|K_3|$.

If $K_1=0$, the gain-bandwidth product is

$$\frac{K_{vD-C}}{\tau'}=\frac{K_2}{T} \tag{10A}$$

Frequency response is easily obtained from the transfer function, equation 6. Gain and phase angle are obtained as the magnitude and the phase angle of $G^*(z)$, equation 6, when $e^{jT\omega}$ is substituted for z in which ω is the input frequency. If the polar co-ordinates are used the frequency response curve is obtained by successive mapping on the complex planes as shown in Fig. 2. The input frequency has been taken only up to the supply voltage frequency because of the limit set by Shannon's sampling theorem.

Self-Saturating Magnetic Amplifier (Alternating Output Circuit) with Multiple-Resistive Control Circuits

As an extension of the magnetic amplifier with a single control circuit, a magnetic amplifier with n-resistive control circuits is analyzed. Although the self-saturating magnetic amplifier (alternating output circuit), whose circuit diagram is given in Fig. 3, is chosen here the same technique can be used for series- or parallel-connected saturable reactors as well as the bridge and the center-tap circuit. With practical assumptions concerning characteristics of cores and rectifiers and modes of operation,[1] the piecewise linear approach[3] can be used for derivation of the difference equation.

When both cores are unsaturated and rectifier no. 1 and no. 2 are blocking and unblocking respectively, the following six equations can be written from the Kirchhoff circuit laws and from the magnetic constraints obtained from the assumption of a rectangular **B-H** loop.

$$e_{cj}=R_{cj}i_c-\frac{e_1}{N_j}+\frac{e_2}{N_j}$$

$$e_s=R_Li_L+R_xi_1+e_1$$

$$e_s=R_Li_L+R_wi_2+e_2$$

$$i_L=i_1+i_2$$

$$\sum_{j=1}^{n}\frac{i_{cj}}{N_j}-i_1=-i_{Om}$$

$$\sum_{j=1}^{n}\frac{i_{cj}}{N_j}+i_2=i_{Om}$$

The current i_{Om} is the magnetizing current (dynamic) referred to the gate winding. Solving these equations simultaneously for e_1, e_2 and i_L with the assumption that $R_x >> R_w$, R_L and $\bar{e}_s >> R_w i_{Gm}$, $R_L i_{Gm}$ yields the following:

$$e_1 = e_s - \frac{R_x}{1+R_x\sum_{j=1}^{n}\frac{1}{N_j^2R_{cj}}}\sum_{j=1}^{n}\frac{e_{cj}}{N_jR_{cj}} - \frac{R_x}{1+R_x\sum_{j=1}^{n}\frac{1}{N_j^2R_{cj}}}i_{Gm} \qquad (11)$$

$$e_2 = e_s \qquad (12)$$

$$i_L = 2i_{Gm} \qquad (13)$$

When core no. 2 is saturated, core no. 1 is on the static loop (i_{Gm}') and rectifier no. 1 and no. 2 are in the same state as in the previous mode, five equations can be written.

$$e_{cj} = R_{cj}i_{cj} - \frac{e_1}{N_j}$$

$$e_s = i_LR_L + R_xi_1 + e_1$$

$$e_s = i_LR_L + R_wi_2$$

$$i_L = i_1 + i_2$$

$$\sum_{j=1}^{n}\frac{i_{cj}}{N_j} - i_1 = -i_{Gm}'$$

Solving for e_1 and i_L with the assumption that $R_x >> R_w$, R_L and $\bar{e}_s >> R_w\bar{i}_{Gm}'$, $R_L\bar{i}_{Gm}'$;

$$e_1 = \frac{\frac{R_w}{R_L+R_w}e_s - R_x\sum_{j=1}^{n}\frac{e_c}{N_jR_{cj}} - R_xi_{Gm}'}{1+R_x\sum_{i=1}^{n}\frac{1}{N_j^2R_{cj}}} \qquad (14)$$

$$i_L = \frac{e_s}{R_L+R_w} \qquad (15)$$

The half-cycle average values of the output voltage during the nth and the $(n+1)$th half-cycles are expressed as follows.

$$\bar{e}_L(n) = \frac{R_L}{R_L+R_w}\bar{e}_s\frac{1+\cos\alpha(n)}{2} \qquad (16)$$

$$\bar{e}_L(n+1) = \frac{R_L}{R_L+R_w}\bar{e}_s\frac{1+\cos\alpha(n+1)}{2} \qquad (17)$$

where the bar above denotes the half-cycle average and α is the firing angle. The flux change of a core during its reset half-cycle is equal to the flux change during the following output half-cycle. This flux condition applied to core no. 1 is expressed in terms of half-cycle average voltage in the following equation.

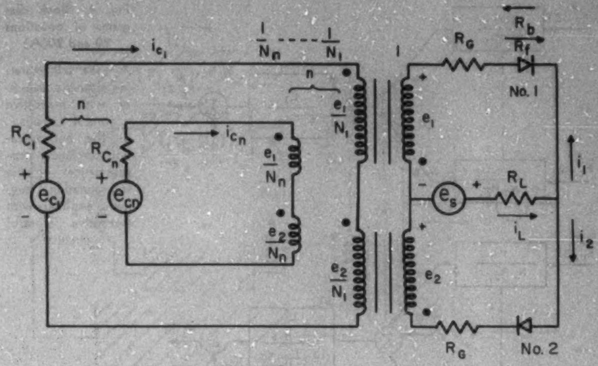


Fig. 3. Circuit diagram of the self-saturating magnetic amplifier (alternating output circuit) with n-resistive control circuits

$$\bar{e}_1(n) = \bar{e}_1(n+1) \qquad (18)$$

With the approximation

$$\frac{\alpha}{\pi} \simeq \frac{1-\cos\alpha}{2}$$

the difference equation is derived from equations 16, 17, and 18 together with equations 11, 12, and 14.

$$\bar{e}_L(n+1) = L_3'\sum_{j=1}^{n}\frac{\bar{e}_{c(n)}}{N_jR_{cj}} + L_3\bar{e}_L(n) + L_4 \qquad (19)$$

where

$$L_3' = \frac{R_L}{R_L+R_w}\frac{R_x}{1+R_x\sum_{j=1}^{n}\frac{1}{N_j^2R_{cj}}}$$

$$L_3 = 1 - \left[1+\frac{R_L+R_w}{R_w}\frac{R_x(\bar{i}_{Gm}-\bar{i}_{Gm}')}{\bar{e}_s}\right]\times \frac{R_w}{R_L+R_w}\frac{1}{1+R_x\sum_{j=1}^{n}\frac{1}{N_j^2R_{cj}}}$$

$$L_4 = \frac{R_L}{R_L+R_w}\frac{R_x}{1+R_x\sum_{j=1}^{n}\frac{1}{N_j^2R_{cj}}}\bar{i}_{Gm}$$

Application of the z-transform method to equation 19 results in the following output equation.

$$E_L^*(z) = \frac{L_3'}{z-L_3}\sum_{j=1}^{n}\frac{E_{cj}^*(z)}{N_jR_{cj}} + \frac{L_4}{1-L_3}\frac{z}{z-1} \qquad (20)$$

The continuous approximation of equation 20 is given as the following

$$E_L(s) = \frac{L_3'}{1-L_3}\frac{1}{1+\frac{1}{1-L_3}Ts}\times \sum_{j=1}^{n}\frac{E_{cj}(s)}{N_jR_{cj}} + \frac{L_4}{1-L_3}\frac{1}{s} \qquad (20A)$$

The block diagram interpretation of equations 20 and 20(A) is given in Fig. 4.

Self-Saturating Magnetic Amplifier (Alternating Output) with *n*-Frequency Sensitive Control Circuits

When a magnetic amplifier is used as a component in a control system, frequency sensitive networks are certain to appear in the control circuits whether they originate in the forward or feedback loop. The interaction of the frequency sensitive networks in the control circuits influences the operation of the magnetic amplifier. The output equation of the magnetic amplifier must therefore be derived in combination with the frequency sensitive networks. The circuit chosen here is the self-saturating magnetic amplifier (alternating output circuit) with n-frequency sensitive control circuits, which are described by $P(s)$ (the ratio between a driving-point impedance and a transfer impedance) and $Q(s)$ (the reciprocal of a transfer admittance) in operational forms. The circuit diagram is given in Fig. 5. Although the alternating output circuit is chosen, the operating characteristics of other circuits, such as a bridge and a center-tap circuit, can be described by the output equation with slight modification of constants.

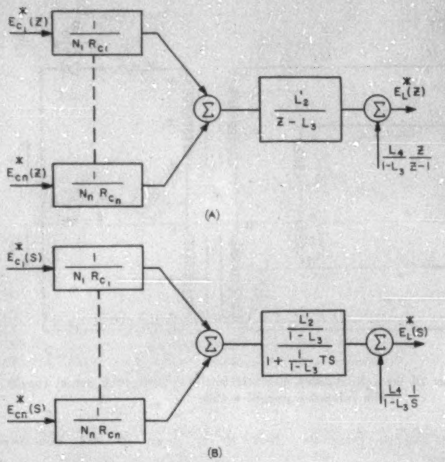


Fig. 4. Block diagrams of equations 20 and 20(A)

A—The self-saturating magnetic amplifier with n-resistive control circuits

B—The approximation continuous representation of the self-saturating magnetic amplifier with n-resistive control circuits

Derivation of the difference equation and the output equation follows exactly the same method as used in the previous section except that all varying quantities are in terms of their Laplace transforms denoted by capital letters. When both cores are unsaturated and rectifier no. 1 and no. 2 are blocking and unblocking respectively, six equations can be written.

$$E_{cj}=P_j\left[-\frac{E_1}{N_j}+\frac{E_2}{N_j}\right]+Q_jI_{cj}$$

$$E_t=R_LI_L+R_xI_1+E_1$$

$$E_s=R_LI_L+R_wI_2+E_2$$

$$\sum_{j=1}^{n}\frac{I_{cj}}{N_j}-I_1=-I_{Gm}$$

$$\sum_{j=1}^{n}\frac{I_{cj}}{N_j}+I_2=I_{Gm}$$

$$I_L=I_1+I_2$$

Solving the equations above simultaneously E_1, E_2, and I_L are obtained

$$E_1=E_s-\frac{R_x}{1+R_x\sum_{j=1}^{n}\frac{P_i}{N_j^2Q_j}}\sum_{j=1}^{n}\frac{E_{cj}}{N_jQ_j}-\frac{R_x}{1+R_x\sum_{i=1}^{n}\frac{P}{N_j^2Q}}I_{Gm} \quad (21)$$

$$E_2=E_s \quad (22)$$

$$I_L=2I_{Gm} \quad (23)$$

When core no. 2 is saturated, core no. 1 is on the static loop (i_{Gm}') and rectifier no. 1 and no. 2 are in the same state as in the previous mode, five equations can be written,

$$E_{cj}=P_j\left(-\frac{E_1}{N_j}\right)+Q_jI_{cj}$$

$$E_s=R_LI_L+R_xI+E_1$$

$$E_s=R_LI_L+R_wI$$

$$\sum_{j=1}^{n}\frac{I_{cj}}{N_j}-I=-I_{Gm}'$$

$$I_L=I+I$$

Solving the equations above simultaneously, E_1 and I_L are obtained as follows,

$$E_1=\frac{1}{1+R_x\sum_{j=1}^{n}\frac{P_j}{N_j^2Q_j}}\left[\frac{R_w}{R_L+R_w}E_s-R_x\sum_{j=1}^{n}\frac{E_{cj}}{N_jQ_j}-R_xI_{Gm}'\right] \quad (24)$$

$$I_L=\frac{E_s}{R_L+R_w} \quad (25)$$

Following exactly the same procedure as in the previous section three determining equations for the difference equation are obtained.

$$\bar{e}_L(n)=\frac{R_L}{R_L+R_w}\bar{e}_s\frac{1+\cos\alpha(n)}{2} \quad (26)$$

$$\bar{e}_L(n+1)=\frac{R_L}{R_L+R_w}\bar{e}_s\frac{1+\cos\alpha(n+1)}{2} \quad (27)$$

$$\bar{e}_1(n)=\bar{e}_1(n+1) \quad (28)$$

The difference equation is derived from equations 26, 27, and 28 together with equations 21, 22, and 24.

$$\bar{e}_L(n+1)=\frac{R_LR_x}{R_L+R_w}I_n\left[\mathfrak{L}^{-1}\left\{F\sum_{j=1}^{n}\times\frac{E_{cj}}{N_jQ_j}\right\}\right]+\left[1-\frac{R_w}{R_L+R_w}k\right]\bar{e}_L(n)+M_4 \quad (29)$$

where

$$k=\left[1+\frac{R_L+R_w}{R_w}\frac{R_x(\bar{i}_{GM}-\bar{i}_{GM})}{\bar{e}_s}\right]\lim_{s\to 0}F$$

$$M_4=\frac{R_LR_x}{R_L+R_w}\bar{i}_{Gm}\lim_{s\to 0}F$$

$$F=\frac{1}{1+R_x\sum_{j=1}^{n}\frac{P_j}{N_j^2Q_j}}$$

$$I_n=\frac{1}{T}\int_{nT}^{(n+1)T}dt$$

Following the same procedure as in the previous section the output equation is obtained

$$E_L^*(z)=\frac{1}{z-M_3}\sum_{j=1}^{n}\left\{Z\left[\frac{M_2'}{n_jQ_j}\right]E_{cj}^*(z)\right\}+\frac{M_4}{1+M_3}\frac{z}{z+1} \quad (30)$$

where

$$M_2'=\frac{R_L}{R_L+R_w}R_xF$$

$$M_3=1-k\frac{R_w}{R_L+R_w}$$

The continuous approximation of equation 30 is

$$E_L(s)=\frac{M_{2a}'}{s+M_{3a}}\sum_{j=1}^{n}\frac{E_{cj}}{N_jQ_j}+\frac{M_4}{1-M_3}\frac{1}{s} \quad (30A)$$

where

$$M_{2a}'=\frac{R_w}{R_L+R_w}\frac{R_x}{T}F$$

$$M_{3a}=\frac{k}{T}\frac{R_w}{R_L+R_w}$$

Block diagram interpretations of equations 30 and 30(A) are given in Fig. 6 and Fig. 6(A).

Simple observation of the block diagram Fig. 6 gives the following conclusions concerning the effects of control circuits on operating characteristics. Inputs are summed up in terms of ampere-turns expressed in operational forms. The response of the output is primarily determined by M_2' and $1/z-M_3$.

The form of M_2' (a rational function

in s) depends only on the configuration of control circuits, but

$$\frac{1}{s-M_3}$$

is the inherent response characteristic of the magnetic amplifier itself, although the constant M_3 is also a function of control circuits. If all control circuits are frequency insensitive the magnetic amplifier becomes a first order amplifier, and its time constant,

$$-\frac{T}{\ln|M_3|}$$

increases as the number of control circuits increases.

It is shown from the following examples that equation 30 is the general characteristic equation which includes all other cases as particular examples.

Case 1. All resistive control circuits:

$$P_j=1,\ Q_j=R_{cj},\ F=\frac{1}{1+R_x\sum_{j=1}^{n}\frac{1}{N_j^2R_{cj}}}$$

Substitution of Q_j and F into equations 30 and 30(A) results in equations 20 and 20(A) respectively.

Case 2. A single resistive control circuit:

$$F=\frac{N^2R_c}{N^2R_c+R_x}$$

Substitution of F into equations 30 and 30(A) results in the output equation of the self-saturating magnetic amplifier.[1] If the continuous approximation is made in this case, with an additional assumption $R_j>>N^2R_c$ the incremental transfer junction becomes the following equation 31

$$G(s)=\frac{K\frac{N_c}{R_c}}{1+\frac{L_c}{R_c}s} \tag{31}$$

in which

$$K=\frac{R_L}{R_w}\frac{1}{1+\frac{R_L+R_w}{R_w}\frac{R_x(\bar{\imath}_{Gm}-\bar{\imath}_{Gm}')}{\bar{e}_s}}\frac{1}{N_G}$$

$$L_c=k_0N_c^2=\frac{\frac{R_L+R_w}{R_w}R_xT}{1+\frac{R_L+R_w}{R_w}\frac{R_x(\bar{\imath}_{Gm}-\bar{\imath}_{Gm}')}{\bar{e}_s}}\times\frac{1}{N_G^2}N_c^2$$

N_G and N_c are numbers of turns of the gate winding and the control winding. Equation 31 may indicate use of the

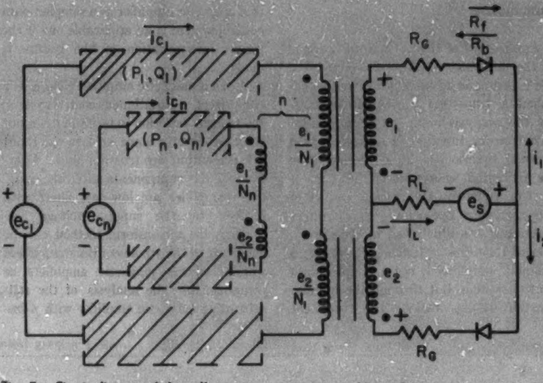


Fig. 5. Circuit diagram of the self-saturating magnetic amplifier (alternating output circuit) with n-frequency sensitive control circuits

analogy to an electric circuit such as a simple R-L circuit.[4]

Case 3. A resistive control circuit and a feedback winding with an impedance:[4]

$$P_j=1$$

$$Q_1=R_c,\ Q_2=Z_f(s)$$

$$F(s)=\frac{1}{1+\frac{R_x}{N_G^2}\left(\frac{N_c^2}{R_c}+\frac{N_f^2}{Z_f(s)}\right)}$$

with

$$R_x>>N_1^2R_c,\ N_2^2Z_f(s)$$

$$F=\frac{N_G^2}{R_x}\frac{1}{\frac{N_c^2}{R_c}+\frac{N_f^2}{Z_f(s)}}$$

Substitution of Q's and $F(s)$ into equation 30(A) with appropriate assumptions results in the following equation which was previously derived.

$$E_L(s)=\frac{K}{1+\left[\frac{L_c}{R_c}+\frac{L_f}{Z_f}\right]s}\times\left[\frac{N_c}{R_c}E_c(s)+\frac{N_f}{Z_f}E_L(s)\right] \tag{32}$$

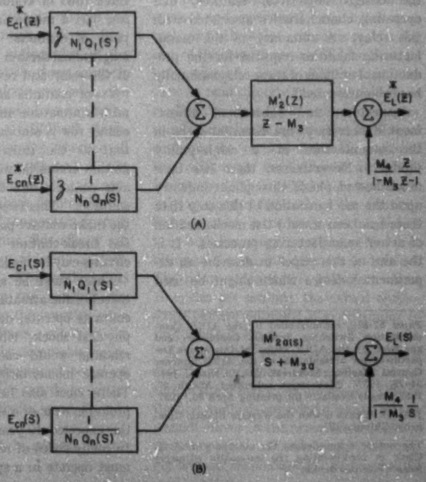


Fig. 6. Block diagrams of equations 30 and 30(A)

A—The self-saturating magnetic amplifier (alternating output circuit) with n-frequency sensitive control circuits

B—The approximate continuous representation of the self-saturating magnetic amplifier (alternating output circuit) with n-frequency sensitive control circuits

Conclusions

Application of the z-transform method to the difference equation has eliminated the continuous approximation and established a consistent mathematical basis for the analysis of magnetic amplifiers. The transfer function is expressed as a rational function of $z=e^{Ts}$ rather than approximated as a rational function of s. Consequently, the discrete nature of the dynamic operation is accurately described. As illustrated in the derivation of the transfer function, however, a magnetic amplifier is not a sampled-data control system, but through the impulse function representation, the treatment of a magnetic amplifier as a sampled-data control system is an applicable one if the proper interpretation of the results is made.

When a magnetic amplifier is used as a component of a continuous control system, the transfer function in terms of z can not be used directly since neither the input nor the output are true impulses. However, if the components after the magnetic amplifier are low-pass filters with respect to the supply voltage frequency, the z-transform method can be extended to the whole continuous system which includes magnetic amplifiers as components. The analysis of the self-saturating magnetic amplifier with n-frequency sensitive control circuits illustrates in detail application of the z-transform method in such cases.

References

1. DYNAMIC OPERATION OF MAGNETIC AMPLIFIERS FOR FEEDBACK CONTROL SYSTEMS, H. C. Bourne, Jr., D. Nitzan. *Scientific Report no. 1, TN-56-555*, Air Force Cambridge Research Center, Cambridge, Mass., May 1956.

2. ANALYSIS AND SYNTHESIS OF SAMPLED-DATA CONTROL SYSTEMS, Eliahu I. Jury. *AIEE Transactions*, vol. 73, pt. I, Sept. 1954, pp. 332–46.

3. MAGNETIC AMPLIFIERS SERVE TO TEACH POWER MODULATORS, H. C. Bourne, R. M. Saunders. *Electrical Engineering*, Oct. 1956, pp. 943–47.

4. ALTERATION OF THE DYNAMIC RESPONSE OF MAGNETIC AMPLIFIERS BY FEEDBACK, R. O. Decker. *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section) pp. 658–65.

Dynamic Gauging and Automatic Adjusting of Wire Spring Relays

JOHN G. WEEKS
ASSOCIATE MEMBER AIEE

RELAYS are the most commonly used electrically operated device in telephone switching systems. A few years ago a radically different design of relay, identified as the "wire spring type," was introduced for use as a basic component of Bell System equipment. Several articles[1,2] have been published that describe the design objectives, features, and operating characteristics associated with this relay. A summary of the manufacturing facilities required for the production of millions of these relays annually has been presented.[3]

Inspection, and subsequent adjustment if necessary, are essential steps in the manufacture of a high-quality product. Nevertheless, there has been much less emphasis throughout industry upon the mechanization of this step than there has been toward the mechanization of actual manufacturing processes. It is the aim of this paper to describe an experimental device which might be used for automatically inspecting, and automatically adjusting if required, some of the operational characteristics of the wire spring relay.

The Problem

A typical wire spring relay, Fig. 1, has three rows of contacts; each individual row has a maximum of 12 contacts or pairs of contacts. Fig. 2 represents a longitudinal section through the front end of the relay and reveals that two of the rows of contacts are movable with the relay's armature and card but that the center row is stationary. It can be seen that as the armature moves relative motion occurs between the rows of movable contacts and the row of stationary contacts. This results in the meeting of the make-contact pairs and the parting of the break-contact pairs; thus electric circuits may be made or broken.

There must be a definite amount of travel of the armature and card after the contacts operate, otherwise the slightest physical shock, vibration, or armature rebound would cause the contacts to operate momentarily when not desired. There must also be some allowance for probable wear of component parts during the expected 40-year life of these relays. In many codes of relays certain contacts must operate in a specified sequence, for example, make before break, and this is ensured by specifying that the make contacts and the break contacts operate within separate ranges of the total armature travel. Furthermore, if all the break contacts of a relay operated very early in the armature's travel the spring load to be counterbalanced by the magnetic pull on the armature would be excessively large for the particular air gap encountered and hence the relay might operate too slowly, or might not meet the minimum operate current requirements. The foregoing may be summarized by stating that if the relay is to perform its function of making and breaking electric circuits rapidly, reliably, and in the proper sequence, it is necessary to specify a range or ranges within the armature's total travel wherein all the contacts must operate. For a typical wire spring relay this range is 0.007 to 0.016 inch of armature travel or follow after the contacts have operated. That is, for each relay contact the contact gauging distance, which is defined as the distance between the armature's position at the instant of contact operation and the armature's position at the end of the operate stroke, must not fall outside the range of 0.007 to 0.016 inch.

A study of the relay design indicates that the contact gauging distances depend upon at least five piece part dimensions. Analysis has indicated that with economically reasonable manufacturing tolerances on these dimensions a certain percentage of the relays will have one or more contact operate points which fall outside of this acceptable range unless adjustments are made to individual relays. Further reference to Fig. 2 shows that the contact operate points can be altered by modifying the position of the

---

Paper 57-810, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 27, 1957; made available for printing April 30, 1957.

JOHN G. WEEKS is with the Western Electric Company, Chicago, Ill.

The author acknowledges the assistance of A. J. Clark in incorporating the automatic adjusting feature into this device.

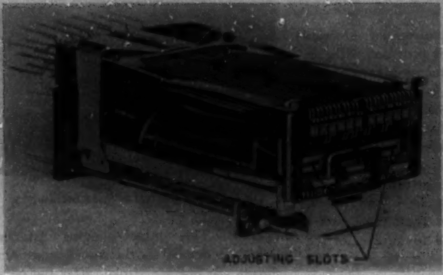


Fig. 1 (above). Wire spring relay

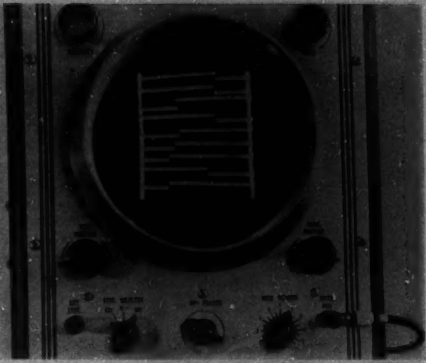


Fig. 4 (right). Oscilloscope showing actual contact gauging pattern

stationary contacts. Fig. 1 shows the provisions that have been made in order that this adjustment may be easily accomplished. It will be noted that a slot is provided in each core plate leg and that these legs hold the center row of contacts in position. A screwdriver blade or adjusting tool may be used to deform the core plate and consequently permanently modify the position of the center row of contacts. It should be emphasized that any adjustment affects not just one contact but all contacts on the relay in varying degrees.

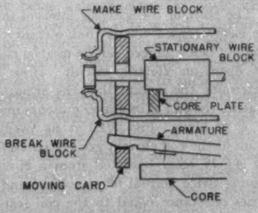


Fig. 2. Longitudinal section through front end of relay

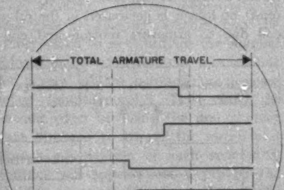


Fig. 3. Oscilloscope contact gauging pattern

Gauging Methods

Having outlined the need for gauging limits and how the contact operate points may be adjusted, it remains to be shown how the contact gauging distances can be measured. Several methods present themselves.

Thickness Gauges

The armature can be displaced from its operated position an amount equal to the minimum acceptable gauging distance, say 0.007 inch, by placing a 0.007 thickness gauge between the core and armature and then operating the relay. All of the the relay's make contacts should be closed and all of the break contacts open; this is ascertained by noting whether the lamps connected through these contacts are lit or not. Conformance with the maximum limit is checked by repeating this procedure using a 0.016-inch thickness gauge and ascertaining whether all the make contacts are open and all the break contacts are closed. If all contacts do not meet these limits then a manual adjustment of the contact operate points is performed and this affects not only the particular out-of-limits contact, but all others as well. This method does not actually measure the gauging distances but indicates to the operator, on a purely "go-no-go" basis, whether the distances are within the acceptable range. Consequently, any adjustment based on this gauging information can do no better than adjust within limits, even if only marginally so.

This is the method in general use today.

Light Probe

In another method considered the actual gauging distances for several contacts are displayed simultaneously thus facilitating the relay adjusting and making it possible to get a more nearly optimum adjustment. (An optimum adjustment would concentrate the contact operate points about the nominal mid-range value.) In this method, an oscilloscope screen is used to display the information in the following manner. The beam is deflected horizontally across the screen in such a manner as to make its position on the screen proportional to the actual physical displacement of the armature. Operation of the contact being gauged causes a vertical step in the sweep, Fig. 3, which is upward for a make contact and downward for a break contact, and when the armature reachs the end of its operate stroke the oscilloscope beam ceases to move. The distance can then be read directly from the calibrated oscilloscope face. In order to display simultaneously the information for several contacts, the vertical deflecting plates of the oscilloscope are switched at a high frequency (100 to 500 kc) between the sensing circuits for each of the contacts being gauged in order to accomplish a very-high-frequency sampling of the condition of the various contacts under test. The relay under test is pulsed at about 10 cycles per second; the actual oscilloscope pattern obtained is very similar to that shown in Fig 4.

A horizontal sweep voltage which is directly proportional to the relay's armature position had to be obtained. One of the first methods considered was to attach a small mirror to the armature; since the armature moves about a hinge the resulting angular displacement can be used to

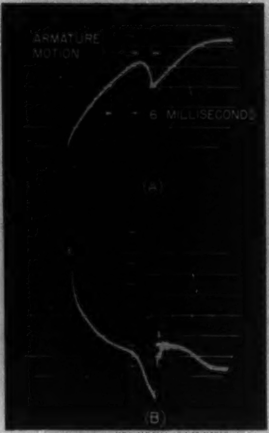


Fig. 5. Oscillograms

A—Coil current growth
B—Differentiated coil current

deflect the reflected light beam across the surface of a phototube. The light-sensitive area is partially masked so as to make the phototube circuit's output voltage proportional to the light beam's position. The use of this output voltage to deflect the oscilloscope beam horizontally across the screen produces the required direct calibrated indication of armature position.

This approach possesses several advantages over the previously described method which uses thickness gauges, such as the following two.

1. Gauging on a variables basis is accomplished instead of on an attributes basis, and consequently the operator's adjusting task is easier and a more nearly optimum adjustment can be obtained.
2. This is a dynamic test, rather than a static test, since the gauging is done with the relay armature actually moving; thus actual operating conditions are more nearly approximated.

However, the following disadvantages must also be weighed:

1. Ambient light variations affect the transducer's calibration and introduce gauging errors.
2. A mirror or other reflecting surface has to be attached to the armature. This is time consuming. The test is not a truly dynamic one since the attachment of mass to the armature changes conditions from those actually encountered in the field.
3. A precise positioning of the relay relative to the sensing system is required.

Mechano-Electronic Transducer

The Radio Corporation of America *5734* electronic tube is provided with an external physical connection to its plate such that when this extension is moved the internal plate position is modified, thus changing the tube's electrical characteristics. It is consequently feasible to use a gauging system similar to the preceding one but to substitute this transducer for the light probe. However, it may be readily seen that this system possesses not only the advantages of the preceding one but also most of the disadvantages.

Capacitance Pickup

The Bell Telephone Laboratories have developed a dynamic gauging system[4] for relays. The over-all system is very similar to the two mentioned previously except that the displacement sensing device does not require physical contact with the relay's armature. One plate of an air capacitor is precisely located relative to the armature; the armature itself serves as the other plate. This capacitor provides part of the capacitance of a resonant circuit; consequently as the armature moves and the capacitance changes so does the resonant frequency of this circuit. This change in resonant frequency is detected and produces a proportional direct voltage output which becomes the oscilloscope's horizontal deflecting signal. Fig. 4 shows the oscilloscope pattern actually obtained when gauging nine contacts simultaneously. Limit lines denoting the two contact ranges could of course be superimposed upon this pattern if desired.

This system performs a truly dynamic test since no physical loading of the relay's armature is used. As a laboratory tool for studying contact behavior during relay operation this device is favored over other approaches. However, it does require the precise positioning of the relay relative to the probe, may require frequent calibration, and is relatively sensitive to stray capacitance effects introduced by the operator's hands or adjusting tools. Consequently, from the standpoint of a production tool, this approach also leaves much to be desired.

Time Interval Measurement

For relays of a given code the physical and electrical characteristics are so nearly identical as to make their armature velocities during operation approximately the same. Furthermore, these velocities are approximately constant during the portion of the travel in which the contacts operate. The contact gauging distance (which is defined as the travel, measured during the operate stroke, after the contact under test operates) can be determined by measuring the time interval between the instant of contact operation and the instant that the armature reaches the end of its travel, and multiplying this by the known armature velocity. This method does not require the precise positioning of the relay under test relative to a sensing device, does not require any physical loading of the armature and is thus a truly dynamic test, and is not subject to error due to ambient conditions as is the case with the previously mentioned dynamic gauging methods. Also, this method requires only that electrical connections be made to the terminals at the back of the relay, thus leaving the front area of the relay entirely free for any adjusting device.

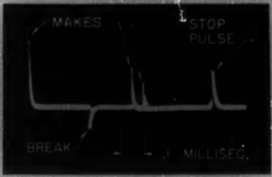


Fig. 6. Time interval gauging oscillogram

The start of the gauging interval may be easily sensed for at that instant the contact being gauged can be used to close or open an electric circuit. How to determine the instant that the armature reaches the end of its travel presents more of a problem. Fig. 5(A) is an oscillogram of the coil current growth curve for an *AF-15* relay operated from a circuit having twice normal relay voltage and a series resistance equal to the coil resistance. It can be seen that when the armature reaches the end of its travel, there is a sharp discontinuity in this curve. This results because a back electromotive force (emf) no longer is generated in the relay coil due to armature motion and con-

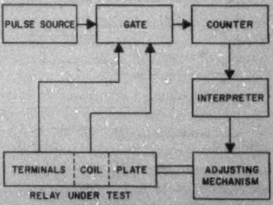


Fig. 7. Dynamic gauging and automatic adjusting schematic

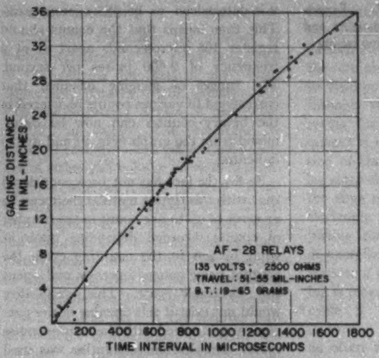


Fig. 8. Gauging distance versus time interval for long-travel relays

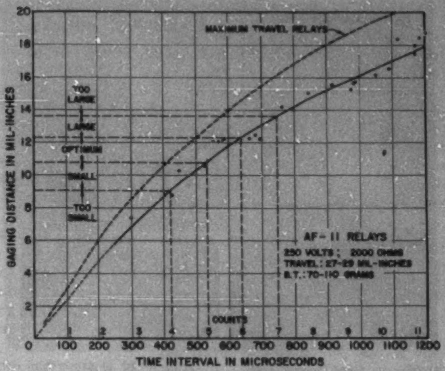


Fig. 9. Gauging distance versus time interval for short-travel relays

sequently the current increases at a more rapid rate. An *R-C* network is used to differentiate this current growth curve. Fig. 5(B) is an oscillogram showing the resulting sharp pulse which coincides with the armature reaching the end of its travel.

This gauging method may be summarized in the following manner: The relay is electrically operated. When the contact being gauged closes or opens the oscilloscope beam is triggered and starts to move horizontally across the tube face at a constant velocity. This sweep speed has been previously set to such a value that actual armature displacements may be read directly off the oscilloscope's calibrated horizontal scale in thousandths of an inch. When the armature reaches the end of its travel this is indicated on the oscilloscope by the pip which results from the differentiated relay coil's current growth curve; thus the length of the sweep line up to the pip is the contact gauging distance and it is read off the scale directly. Several refinements of this general gauging scheme may be incorporated so as to provide the gauging information for not only one contact, but for several of the contacts on a given relay simultaneously. This could be accomplished by using the first contact that operates to trigger the sweep and then the operation of each succeeding contact will superimpose a pip on the oscilloscope sweep to indicate the point in the armature's travel at which it operates. The resulting oscilloscope pattern is shown in Fig. 6 for an *AF-15* relay operated from 60 volts through 5,000 ohms. In order to identify each contact being gauged, the height of its associated pip could be varied in proportion to its position on the relay, i.e., the extreme left-end contact would produce a tall pip and the extreme right-end contact would produce a short pip with contacts in between these extremes producing pips of varying intermediate heights. Also, make contacts could be identified by positive pips and break contacts by negative pips.

With the method outlined it is still necessary for an operator to interpret the information shown on the oscilloscope and to perform the adjustment. It appeared feasible to proceed to a more automatic gauging and adjusting scheme and eliminate, or at least reduce, the effort of an operator. Fig. 7 is a simplified block diagram of the scheme developed. The relay under test is electrically operated. The operation of the individual contact being gauged opens an electronic gate and allows pulses to pass through the gate to the electronic counter until, some time

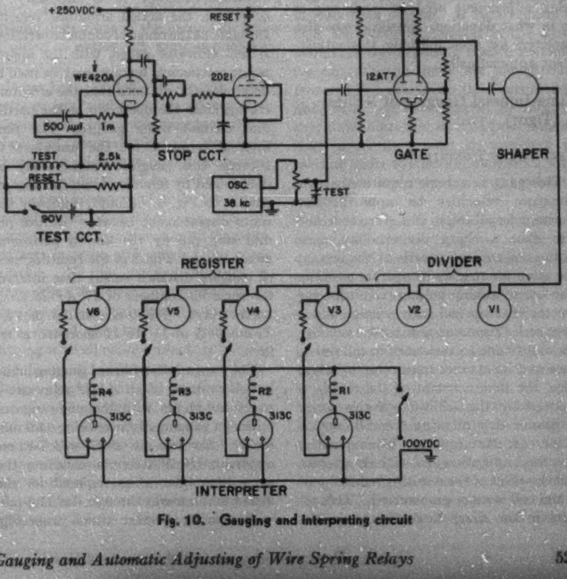


Fig. 10. Gauging and interpreting circuit

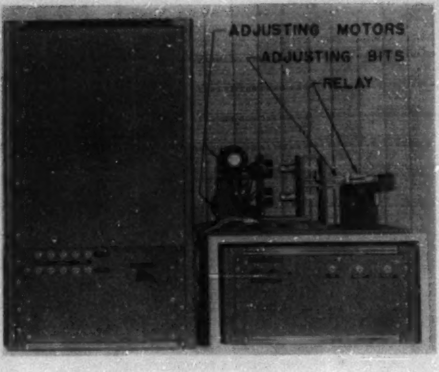


Fig. 11. Experimental gauging and adjusting facility

later, the armature reaches the end of its travel and the gate is closed by the stop pulse from the differentiated current growth curve. The number of pulses registered is a measure of the time required for the armature to travel through the gauging distance and, therefore, if the correct pulse source frequency is used this number of counts is a direct indication of the actual gauging distance. The interpreting unit compares the number of counts actually registered with the number of counts that would have been registered had the gauging distance been within established limits and subsequently tells the adjusting mechanism whether a relay adjustment need be made and if so in what direction. Facilities are also depicted for accomplishing this adjustment automatically.

Time Interval Gauging of Relays (Tigor)

Test Relay Conditions

This gauging scheme requires that the armature velocities be approximately constant for all relays of a given code and also that a sharp discontinuity exist in the coil current growth at the instant the armature reaches its operate position. The use of a relay pulsing circuit having greater than normal relay voltage and a large series resistance reduces the velocity variability due to variations in coil resistance and inductance from relay to relay. Also, the time constant of the circuit is reduced by the addition of this series resistance thus allowing the coil current to increase more rapidly. Consequently, the final armature velocity is greater and the effect of the resulting induced emf in the coil is more pronounced. This results in the sharp discontinuity in the coil current growth curve as Figs. 5(A) and 5(B) show. It is interesting to note that these oscillograms clearly show that the armature rebounds and vibrates under the test conditions being used here. Throughout these tests the applied voltage and series resistance have been chosen so as to provide a relay coil current which is slightly more than the current that is normally used in actual field operations.

Gauging Details

In order to determine the frequency of the pulses that are to be gated to the electronic counter it is necessary to determine the actual armature velocity throughout the range of contact operation. A dial indicator gauge, with the appropriate mechanical linkage, may be used to determine the distance that the armature travels after the operation of the particular contact being gauged. The time interval required for the armature to traverse this gauging distance may be determined by referring to an oscillogram similar to Fig. 6. (Simply multiply the units displacement between operate pip and stop pip by the known horizontal sweep speed.) Fig. 8 is the resulting plot of gauging distance versus time interval for some 62 contacts of 17 *AF-28* long-travel relays. Fig. 9 is a similar plot for 24 contacts of 11 *AF-11* short-travel relays.

The presently established gauging limits for all contacts of an *AF-11* relay are 7 to 16 mil-inches with the mid-range or optimum gauging distance being 11.5 mil-inches. For reasons which will become apparent, let $5^{1}/_{2}$ counts consume the same time interval as required for the armature to travel through the 11.5-mil-inch gauging distance which from Fig. 9 is determined to be 585 microseconds. This then means that the counts should register 106 microseconds apart or at a frequency of 9,400 pulses per second. The ranges of gauging distance that correspond to various counts registered in the binary counter can now be determined from the graph (Fig. 9) and are so indicated.

So far the fact has not been considered that with this timing method there can be an error of up to one count in the number of counts recorded depending upon at what instant in the counting cycle the contact being gauged operates and opens the electronic gate. This variability would not exist if this opening of the gate were synchronized with the counter pulse cycle. An alternative solution was used here. The oscillator frequency was increased by a factor of four and consequently the per-cent variability in the time interval measurement was reduced by a like factor. This means, however, that the capacity of the binary counter must now be increased by a factor of four which requires the addition of two more stages.

Fig. 10 is a circuit diagram of the complete system. The test relay is electrically operated and when the break contact being gauged operates, oscillator pulses, at a frequency of 38 kc, are fed into the *12AT7* gating circuit. Simultaneous with the operation of the test relay the reset relay is energized, but since it is a much faster relay its make contact in the plate of the thyratron stop circuit operates before the contact being gauged operates. The bias on the *2D21* thyraton is adjusted so that the tube does not conduct and consequently the *12AT7* allows amplified oscillator pulses, beginning at the instant that the test relay contact operates, to be passed on to a pulse shaper. When the test relay reaches the end of its travel the differentiating circuit stop pulse is amplified and triggers the *2D21* thyratron which subsequently reduces the voltage on one of the grids of the *12AT7* thus blocking the passage of subsequent pulses into the counter chain. The actual electronic counter is comprised of commercially available bistable flip-flops. The first two flip-flops in the chain act only as a divider and the last four comprise the actual pulse register. Note also that the last counter in the chain is so connected that once having flipped it will not flop back unless it is reset.

This eliminates the possibility of ambiguity for a gauging distance so large as to correspond to a time interval in excess of the counter capacity of 16 counts.

Interpreting Unit

This unit translates the number of counts registered into a visual indication of the gauging information and, more important, dictates to the adjusting device the adjusting action, if any, that should be taken. The counters or flip-flops each provide a plate output voltage of either 135 volts or 55 volts depending upon whether it is in the reset "zero" condition or in the flipped "one" condition, respectively. After the test relay has operated, and hence the timing completed, each counter's output is connected to the control electrode of its associated cold-cathode gas tube, Western Electric *313C;* see Fig. 10. If the counter output voltage is 135 volts, the *313C* will conduct and the relay in its anode circuit will be energized. Conversely, if the counter's output voltage is 55 volts the *313C* will not conduct and the relay will not be energized. The four relays are connected to form a modified relay tree such that the total gauging range is split into five cells as shown in Table I.

Table I. Gauging Range

Counts	Range	Indication	Adjustment
<4	<9.0	Too small	Increase
4	9.0 to 10.7	Small	Increase
5	10.7 to 12.3	*Optimum*	*None*
6	12.3 to 13.6	Large	Decrease
>6	>13.6	Too large	Decrease

Adjusting Mechanism

Little consideration has thus far been given to the fact that each relay may possess not only one contact but up to 24 contacts, and that any adjustment affects all of them in varying degrees. However, since the rows of contacts are straight, within close tolerances, if the extreme end contacts are adjusted to mid-range all the other contacts will likewise be close to mid-range or optimum. As a consequence of this reasoning, only the two extreme end contacts are inspected during the initial gauging and adjusting cycle.

Fig. 11 shows the experimental gauging and adjusting device. Each synchronous adjusting motor is energized to rotate in the direction dictated by the interpreting unit and associated relay tree. The torque is transmitted through a magnetic particle fluid clutch and gear train to the adjusting bit which has previously been inserted in the core plate slot. Initially a small current is passed through the clutch, and so only a low torque is transmitted to the adjusting bits. This low torque takes up any play in the various mechanical members, and stresses the core plate leg to some point just below its yield point. Subsequently the current through the clutch is increased to full rating for a predetermined time; this assures a definite known angular deflection of the adjusting bit and core plate leg since the motor speed is constant and the time of rotation is controlled by an electronic timer. At the end of this adjusting interval the motor and clutch are de-energized and the core plate springs back to a new position. The relay gauging is repeated and likewise the adjusting cycle if the gauging information so dictates. This complete cycle is repeated as often as required, within reasonable limits, to bring the two end contacts to the optimum adjustment. The complete cycle of gauging and adjusting is under the control of a motor-driven sequence timer. The gauging for the extreme upper contact is performed and stored in relays, then the extreme lower contact is gauged, and subsequently the upper and lower adjusting bits simultaneously perform the adjustments that are called for. If this 3-second cycle is repeated three times without any appreciable shift in the gauging of the contacts (from "Too Large" to "Large" for instance) the cycles are next performed with longer adjusting time intervals. This refinement is required since the physical characteristics of the core plates do vary one from the other, and this variation necessitates varying deflections of the core plate legs in order to obtain the same permanent deformation.

It has been pointed out that this automatic gauging and adjusting scheme considers only the extreme end contacts of the relay and assumes that if these are adjusted to mid-range or optimum then all other contacts will be within limits. In practice, however, all of the remaining contacts must be checked also; this can be readily done by connecting all the make contacts in series and gauging this combination. If this gauging distance is greater than the lower gauging limit, it has been determined that all of the make contacts exceed this limit. A subsequent gauging with all make contacts in parallel may be used to check the upper limit. Similar gaugings with all break contacts in series and parallel will check the upper and lower limits, respectively. If any one of the contacts did not meet those limits which were established as acceptable, but not necessarily optimum, the relay would be set aside for an individual adjustment of the misaligned contact.

Special Considerations

The basis for the method of gauging discussed assumes that the displacement time relationships for all relays of a given code are very nearly identical. Two major relay variables, total armature travel and back tension, affect this relationship appreciably; furthermore, both of these factors are not closely controlled. The total armature travel for short-travel relays has limits of 26 ±5 mil-inches. The dotted curve on Fig. 9 indicates the effect of increased total armature travel. If the same time interval limits are used, it can be seen that this has the effect of raising the mid-range optimum value and raising the upper limit several mil-inches. However, this may be permissible since if the travel is greater by, say, 5 mil-inches and the upper limit is raised only 3 mil-inches, then the allowance for wear and rebound is actually increased, not decreased. The major consideration is to ensure that such a modification of limits will not appreciably slow down the relay or increase the minimum relay operate current above the desired value. This automatic modification of displacement gauging limits for a given code depending upon total armature travel appears to have the advantage that fewer relays would require individual contact adjustment. Where the total travel is minimum the limits would be pulled in and conversely, for travels near the maximum, the limits would be moved out.

A similar reasoning applies to variations in relay back tension. If back tension is high the curve is modified in the same manner as by increasing total travel and consequently the limits are moved out. however, this also seems to be permissible since with large back tension the effect of shock and vibration would be minimized. Once again care must be exercised in revising limits to ensure conformance with relay speed and operate current requirements.

For relay codes having long travels consideration must be given to the sequencing of contact operations such as make before break. This sequencing is inherent in the relay design by virtue of the shape of the card that moves the contacts. This means that if contacts in one range are automatically gauged and adjusted, then all that is necessary is to check the other contacts for conformance with the limits of the other ranges. This can be done by using a series parallel arrangement of these contacts, similar to that already explained, except that the oscillator frequency would be changed to correspond to the limits that apply to these contacts. All switching associated with these various tests can, of course, be automatically sequenced.

Summary

This paper has outlined a new experimental approach to automatic relay gauging and adjusting. This new concept might well result in an end product that is adjusted more nearly to optimum. If it also ensues that as a result of this method the allowable contact operate range can be economically reduced, a faster operating relay could be designed with attendant over-all system advantages.

A different approach to the measurement of dimensional characteristics has been outlined. Relay gauging is obviously only one of many applications to which this principle may be applied advantageously.

References

1. A New General-Purpose Relay for Telephone Switching Systems, A. C. Keller. *AIEE Transactions*, vol. 71, pt. I, 1952 (Jan. 1953 section), pp. 413–28.

2. Design Features of Bell System Wire Spring Relays, H. M. Knapp. *Ibid.*, vol. 75, pt. I, Sept. 1956, pp. 482–86.

3. Manufacture of Wire Spring Relays for Communication Switching Systems, James W. Rice. *Ibid.*, Nov. 1956, pp. 513–18.

4. Electronic Relay Tester, T. E. Davis, A. L. Blaha. *Bell System Technical Journal*, New York, N. Y., July 1954, pp. 925–38.

High-Frequency Magnetic Servo Amplifier for Air-Borne Control Systems

W. R. SEEGMILLER
ASSOCIATE MEMBER AIEE

THE ADVANTAGES of operating magnetic amplifiers at frequencies higher than the usual 60 and 400 cycles per second (cps) are improved transient response and reduced size and weight. These advantages are very important in the application of magnetic amplifiers to air-borne control systems. This paper presents results of the development of a high-frequency 4,700-cps magnetic servo amplifier used to drive a 400-cps servo motor for military aircraft applications.

Other papers[1,2,3] have been presented on the use of transistor oscillators as a supply for high-frequency magnetic amplifiers. However, the ambient temperature requirements of many military aircraft applications precludes the use of germanium transistors. Also the lack of silicon transistors with sufficient power handling capability has limited silicon-transistor oscillator power supplies to low-level preamplifier stages. Recent developments in small high-frequency generators for aircraft and missile applications have provided a source of high-frequency power which is suitable for driving the output stages of magnetic-amplifier control systems.

A motor-generator set was used as a high-frequency supply for this work. The set consists of a 3-phase 400-cps induction motor and a high-frequency generator mounted on the same shaft. The generator frequency is approximately 4,700 cps. (Synchronous frequency would be 4,800 cps.) This machine weighs $5^3/_4$ pounds and has a rated output of 250 watts at 4,700 cps.

In some air-borne control systems it will be possible to accomplish an over-all reduction in system weight by using this type of machine to convert the 400-cps power to high frequency. In a large complex control system the reduction in transformer and reactor weights can more than offset the added weight of the motor-generator set.

In other control systems where a small engine-driven generator must be furnished with the control (as in many jet-engine control systems, for example) the reduction in over-all system weight by changing to high-frequency operation is even more advantageous. In this case there is a weight reduction on both the generator and the control amplifiers. The servo amplifier application reported on in this paper is part of a general development program of high-frequency magnetic amplifiers for air-borne control systems.

Description of Servo Amplifier

A circuit diagram of the amplifier is shown in Fig. 1. The amplifier consists of three half-wave bridge stages[4] in cascade designed for operation at the 4,700-cps supply frequency. Each stage consists of two toroidal cores of 1-mil Hy-Mu 80 tape. Core and winding data on each stage are given in Tables I and II. Silicon diodes are used for all the rectifiers. A resistance of 100,000 ohms is shunted across each diode to set the operating bias point of the stage.

The output of the amplifier drives a 400-cps two-phase servo motor. The motor used in this application is the BuOrd Mk. 14 size which requires 3.5 watts of control phase power for rated torque of 0.63 inch ounces. The control phase of the motor is wound for a rated voltage of 28.7 volts. Fig. 2 shows the application of this amplifier in a typical position servo loop using 400-cps synchros for the position error signal and a 400-cps tachometer (integral with the servo motor) for rate feedback.

Operation of the amplifier at a frequency higher than the motor frequency eliminates the effect of the usual drift error present in conventional magnetic-amplifier systems.[5] However, in systems requiring high amplifier gain there is still a problem since the d-c drift in the first stage may become large enough to drive the output stage into saturation. One solution to this problem which has been used on full-wave stages is to capacitively couple[3] each stage. In this application of half-wave stages each stage is coupled directly, and a feedback network is used to compensate for any d-c unbalance or drift in the amplifier.

The feedback network, consisting of *R4*, *R5*, and *C1*, is shown in Fig. 2. The purpose of this network is to provide a feedback proportional to any d-c unbalance or drift within the amplifier. Without this network the unbalance or drift of the first and second stages of the amplifier may be sufficient to drive the output stage into saturation. The network keeps the amplifier in the linear region of operation by providing

---

Paper 57-775, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted February 15, 1957; made available for printing April 24, 1957.

W. R. Seegmiller is with the General Electric Company, Schenectady, N. Y.

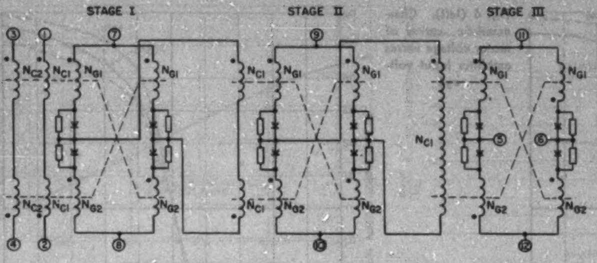


Fig. 1. Circuit diagram of 3-stage half-wave magnetic amplifier. Dotted lines indicate windings on same core

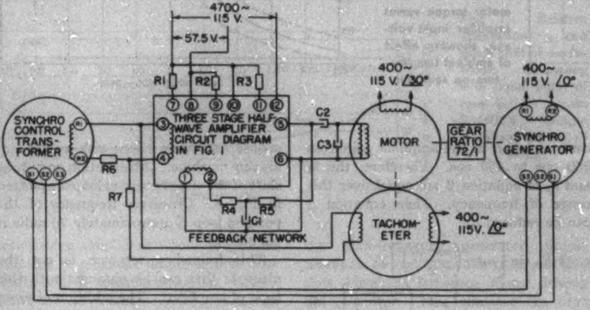


Fig. 2. Schematic diagram of position servo

R1, R2, R3 = 200 ohms
R4, R5 = 24,000 ohms
R6, R7 = 10,000 ohms

C1 = 2.5 microfarads
C2 = 5 microfarads
C3 = 2 microfarads

a feed-back proportional to the d-c unbalance or drift. Capacitor *C1* presents a low impedance to 400 cps, so the feedback network does not appreciably affect the 400-cps gain. Capacitor *C2* is a blocking capacitor which keeps any d-c output off the motor winding. Capacitor *C3* is the usual tuning capacitor chosen for parallel resonance at 400 cps. Capacitor *C3* also acts as a bypass for the 4,700-cps component to keep it off the motor winding.

The error due to drift has been a limiting factor in many magnetic-amplifier applications, particularly on systems requiring high accuracy over a wide range of ambient temperature conditions. The elimination of the drift problem is another important advantage to be gained from operating high-frequency magnetic amplifiers in this manner.

A packaged amplifier is shown in Fig. 3. The can is $1^1/_2$ inches square and $1^1/_4$ inches high. Weight of the amplifier package, after vacuum impregnation and plotting, is only 5 ounces. The total weight, including external resistors and capacitors, is 8 ounces.

Performance Characteristics

Gain Curves and Ambient Temperature Effects

Curves showing motor voltage and stalled motor torque as a function of input voltage to the amplifier are plotted in Figs. 4 and 5. These characteristic curves were taken with the feedback network, blocking capacitor, and tuning capacitor connected as shown in Fig. 2. The tachometer and synchros were disconnected, and a 400-cps voltage source used for the input signal. The input resistance, as shown in Fig. 2, is 10,000 ohms.

The effect of ambient temperature extremes (−55 C (degrees centigrade) and +100 C) on the amplifier gain is shown in Fig. 5. The curves indicate that the amplifier gain in the linear region increases 24% at −55 C and decreases 18% at +100 C. This gain change is due to the combined effect of the extreme ambient temperature on the three stages in cascade. Gain variations of this magnitude are not objectionable for most servo amplifier applications. The data plotted in Fig. 5 also show that there is no zero drift on a servo amplifier used in this manner.

Phase Shift Variation with Signal Frequency

The phase shift between the amplifier input voltage and the output voltage on the motor was measured as the signal frequency was varied between 350 cps and 450 cps. These data are plotted in Fig. 6. The nominal phase shift is 35 degrees lagging at 400 cps. There is a variation of ±10 degrees over the normal operating region of 380 to 420 cps. The circuit used for these data is as shown in Fig. 2, but with tachometer and synchros disconnected and the motor stalled.

The data plotted in Fig. 6 were measured at an input level within the linear range of amplifier gain. This phase shift increases as the amplifier approaches maximum output. Therefore, the fixed phase excitation for the motor was set at 30 degrees leading (see Fig. 2) as a compromise which gave good torque characteristics over the full range of input voltage.

Attenuation and Phase Shift with Modulated Signal

To determine the response of the amplifier and associated networks to a modulated 400 cps signal of the form

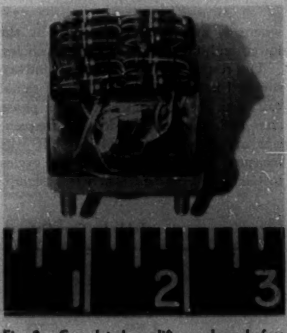


Fig. 3. Completed amplifier package before potting

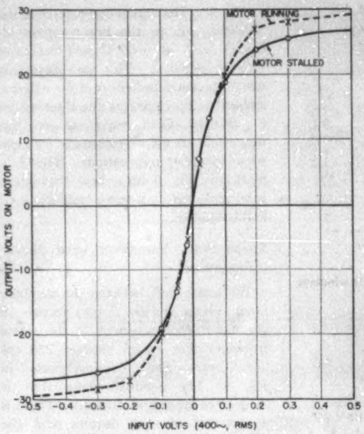


Fig. 4 (left). Characteristic curves of motor voltage versus amplifier input voltage

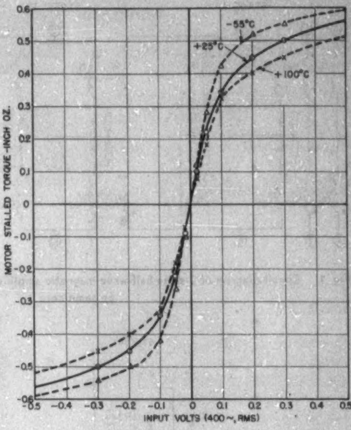


Fig. 5. Characteristic curves of stalled motor torque versus amplifier input voltage, showing effect of ambient temperature on amplifier

$(E_m \sin \omega_s t) \sin (2\pi)(400)t$, the following relations can be used:

$$E_m \sin \omega_s t \sin \omega_c t = \frac{E_m}{2} \cos (\omega_c - \omega_s)t - \frac{E_m}{2} \cos (\omega_c + \omega_s)t \quad (1)$$

where ω_c is the carrier frequency and ω_s is the modulating frequency.

Let α_1 be the phase shift of the amplifier at the lower side-band frequency, $(\omega_c - \omega_s)$, and α_2 be the phase shift of the amplifier at the upper side-band frequency, $(\omega_c + \omega_s)$. Then the output of the amplifier will be of the form:

$$e_o = k_1 \frac{E_m}{2} \cos [(\omega_c - \omega_s)t - \alpha_1] - k_2 \frac{E_m}{2} \cos [(\omega_c + \omega_s)t - \alpha_2] \quad (2)$$

where k_1 and k_2 are the gains at the lower and upper side-band frequencies.

The response of the amplifier and associated networks to a modulated signal can then be determined if the phase shift and gain at the side-band frequencies are known. The phase shift was measured over the range from 350 cycles to 450 cycles (Fig. 6). There is no significant variation in the amplifier gain from 350 cps to 450 cps. Therefore, the k_1 and k_2 of equation 2 are equal over this range of frequency. Then equation 2 can be reduced to:

$$e_o = k_1 E_m \sin \left[\omega_s t - \left(\frac{\alpha_2 - \alpha_1}{2} \right) \right] \times \sin \left[\omega_c t - \left(\frac{\alpha_2 + \alpha_1}{2} \right) \right] \quad (3)$$

Equation 3 indicates that the phase shift of the modulated frequency will be one-half the difference of the phase shifts of the two side-band frequencies.

The phase shift of the modulating frequency is plotted in Fig. 7. These data were calculated from the data in Fig. 6. An experimental check was also made by comparing the envelopes of the modulated waves of input and output of a dual-beam oscilloscope. These data checked that given in Fig. 7 within five degrees. The limit of accuracy to be expected in comparing the envelopes of the modulated waves is about 5 degrees.

Servo Frequency Response

A typical position servo was set up as shown in Fig. 2. The synchro gradient is 1 volt per degree. Frequency response data were taken with a servo analyzing machine. Attenuation and phase shift of the closed position loop are plotted in Fig. 8. Crossover frequency of the position loop is approximately 70 radians per second.

This frequency response is not the ultimate that can be obtained with this type of amplifier. The addition of more stages of gain would make it possible to extend the frequency response well beyond that shown in Fig. 8. This is evident from an examination of the amplifier phase-shift data shown in Fig. 7 with a modulated 400-cps signal. There is little phase shift to beyond 300 radians per second. Practical considerations, such as amplifier saturation due to quadrature signals and harmonics from the tachometer and synchros, are more likely to limit the ultimate frequency response that can be obtained with this type of amplifier than the amplifier phase shift. This is quite different from conventional 400-cps mag-

Table I. Core Data on Magnetic Amplifier Stages

Stage No.	Outer Diameter, Inches	Inner Diameter, Inches	Tape Width, Inches	Tape Thickness, Inches	Material
I	0.500	0.375	0.125	0.001	Hy-Mu 80
II	0.500	0.375	0.125	0.001	Hy-Mu 80
III	0.750	0.500	0.125	0.001	Hy-Mu 80

Table II. Winding Data on Magnetic Amplifier Stages

Stage No.	Winding	Turns	Wire Size AWG No.	Resistance, ohms
I	N_{G1}	600	42	63
	N_{G2}	600	42	65
	N_{C1}	100	42	14
	N_{C2}	200	42	26
II	N_{G1}	600	40	42
	N_{G2}	600	40	48
	N_{C1}	20	40	3
III	N_{G1}	600	30	20
	N_{G2}	600	36	23
	N_{C1}	20	36	2

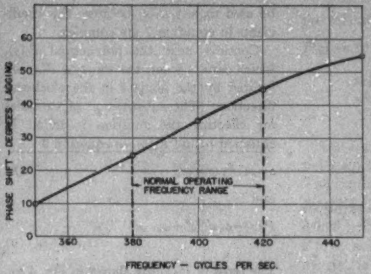


Fig. 6. Phase shift between amplifier input voltage and motor voltage as a function of frequency

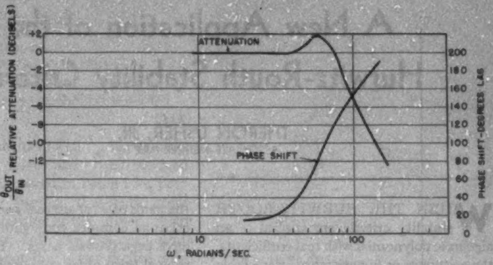


Fig. 8. Frequency response of position servo

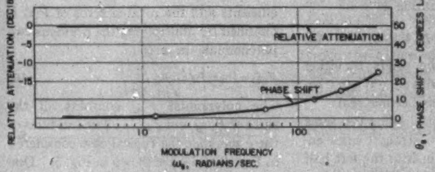


Fig. 7. Relative attenuation and phase shift of modulation frequency with signal input of form $E_m \sin \omega_s t \sin 2\pi(400)t$. Output is voltage on stalled motor

netic servo amplifiers where the phase shift due to magnetic-amplifier time constants and/or transportation lag usually limits the servo frequency response that can be obtained.

Low-Frequency Beat

With the position servo set up as shown in Fig. 2 a low-frequency oscillation was observed at the motor output shaft. The frequency of this oscillation corresponds to the slip frequency of the 400-cps induction motor driving the high-frequency alternator. The magnitude of the oscillation is only a few degrees of rotation of the motor shaft. The oscillation at the synchros output is reduced by the 72 to 1 gear ratio.

An investigation of the cause of the oscillation showed that there is a voltage output at null of about 2 volts across the motor with the tachometer and synchros disconnected. When this null voltage was compared with the 400-cps line it was observed that it beat with the line. A resistor was placed across the high-frequency generator output terminals to load down the generator and slow down the induction motor driving the generator. This increased the beat frequency. The frequency of the oscillation was thus shown to correspond to the slip of the 400-cps induction motor driving the generator.

It should be noted that in practical systems where the 400-cps and 4,800-cps generators are both geared to the same shaft there would be no problem with a beat frequency. Or if the induction motor drive is to be used, a machine with a higher slip frequency could be used to increase the beat frequency high enough that the servo will not follow it. Actually, the magnitude of the oscillation, even with the present motor-generator, is small enough so that it would not be detrimental on many systems.

Conclusions

The advantages of high-frequency operation of magnetic amplifiers in size and weight reduction, and in improved performance, are illustrated in this application.

The usual drift error, inherent in conventional magnetic servo amplifiers, has been eliminated by operating the amplifier at a higher frequency than the motor frequency. The addition of a feedback network which responds to any d-c unbalance in the output makes it possible to cascade several stages without the danger of saturating the output stage as a result of input stage drift.

References

1. A Switching Transistor D-C to A-C Converter Having an Output Frequency Proportional to the D-C Input Voltage, G. H. Royer. *AIEE Transactions*, vol. 74, pt. I, July 1955, pp. 322–26.

2. Operational Magnetic Amplifier with Audio-Frequency Transistor Power Supply, R. O. Decker, Francis Gourash. *Ibid.*, Sept. 1955, pp. 490–96.

3. Capacitively Coupled Magnetic Amplifiers, H. William Collins. *AIEE Special Publication T-86*, "Conference on Magnetic Amplifiers," July 1956, pp. 211–20.

4. Design Considerations of the Half-Wave Bridge Magnetic Amplifier, C. W. Lufcy, H. H. Woodson. *AIEE Transactions*, vol. 73, pt. I, July 1954, pp. 220–25.

5. Elimination of Asymmetry Zero-Drift Errors in Magnetic Servo Amplifiers, W. A. Geyger. *Ibid.*, vol. 74, pt. I, May 1955, pp. 239–45.

6. An All-Magnetic Audio Amplifier System, J. J. Suozzi, E. T. Hopper. *Ibid.*, July 1955, pp. 297–301.

7. Decicycle Magnetic Amplifier Systems for Servo Applications, Leo J. Johnson, S. E. Rauch. *Ibid.*, Nov. 1955, pp. 667–72.

8. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., Chapman and Hall, Ltd., London, England, 1955.

A New Application of the Hurwitz-Routh Stability Criteria

THERON USHER, JR.
ASSOCIATE MEMBER AIEE

WHEN THE HURWITZ-ROUTH stability criteria are applied to an algebraic polynomial with real coefficients, the existence, or nonexistence, of unstable roots of the polynomial is revealed. However, if the polynomial contains all stable roots, the criteria do not give an indication of the relative stability of the roots. The latter information is frequently desired.

For instance, in servomechanism practice, a minimum allowable damping ratio for the roots of the characteristic equation of a system is often specified. For most applications, it is not sufficient to know whether the system is merely stable or unstable. The following method may be used to detect roots whose damping ratios are less than a certain specified value. The usefulness of the Hurwitz-Routh criteria is thus enhanced. A numerical example is also given.

Mathematical Development

The characteristic equation of a physical system is usually represented by the nth order polynomial in equation 1. The coefficients are real.

$$P(s)=a_ns^n+a_{n-1}s^{n-1}+\ldots+a_2s^2+a_1s+a_0 \quad (1)$$

It is assumed that the necessary, but not sufficient, conditions for stability are satisfied:

$$a_n>0(n\geq 1); \quad a_0\geq 0 \quad (2)$$

The Hurwitz-Routh criteria are then applied directly to the coefficients of $P(s)$ to determine if all the zeros lie in the left half of the s-plane. If the roots are all stable, then the $(n\text{-}1)$th order determinant of the Hurwitz array of the coefficients is positive:

$$\begin{vmatrix} a_1 & a_0 & 0 & 0\ldots 0 \\ a_3 & a_2 & a_1 & a_0\ldots 0 \\ a_5 & a_4 & a_3 & a_2\ldots 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0\ldots a_{n-1} \end{vmatrix} >0 \quad (3)$$

For the remainder of the development, the assumption will also be made that the coefficients of $P(s)$ satisfy the inequality in equation 3.

Since the coefficients of $P(s)$ are real, the zeros of $P(s)$ are real and negative, or occur in complex conjugate pairs, which have negative real parts. The zeros of a typical third-order polynomial are shown in Fig. 1. The polynomial contains one real zero and a complex conjugate pair of zeros.

In terms of the relative damping factor, ζ, a pair of complex conjugate zeros may be represented by equation 4.

$$s=\omega_0(-\zeta\pm j\sqrt{1-\zeta^2}) \quad \zeta<1 \quad (4)$$

The distance from the origin to either complex zero so defined is ω_0. When ζ becomes unity, both zeros become real.

The locus of all roots having the same damping factor is two straight lines extending from the origin into the left half half s-plane, and making equal angles with both halves of the imaginary axis. The relation between this angle and the damping factor is found by consideration of the triangle in Fig. 2. This relation is given in equation 5.

$$\cos\theta=\sqrt{1-\zeta^2} \quad (5)$$

Therefore any roots found in the shaded sector of Fig. 2 have a damping ratio greater than the value specified by equation 5.

The problem now is to detect the presence of any zeros of $P(s)$ which lie outside the shaded sector in Fig. 2, but which still remain in the left-half s-plane. The problem is attacked by transforming $P(s)$ into a new polynomial whose right-half-plane zeros correspond to the zeros of $P(s)$ which are located in the sectors just mentioned. The Hurwitz determinant can then be used to detect the unstable roots of the new polynomial.

The roots of $P(s)$ are rotated by the angle θ in the counterclockwise direction when the change of variables indicated in equation 6 is made.

$$z=s\epsilon^{j\theta}; \quad s=z\epsilon^{-j\theta} \quad (6)$$

The corresponding polynomial in z becomes

$$P_1(z)=a_n\epsilon^{-jn\theta}z^n+a_{n-1}\epsilon^{-j(n-1)\theta}z^{n-1}+\ldots+a_1\epsilon^{-j\theta}z+a_0 \quad (7)$$

The zeros of $P_1(z)$ are shown in Fig. 3 for the typical third-order polynomial in Fig. 1. The Hurwitz determinant cannot be used to test $P_1(z)$ because the coefficients in equation 7 are complex.

Consider next the polynomial $P_2(z)$, which results when the zeros of $P(s)$ are rotated by the angle θ in the clockwise direction. The substitution of variables for effecting the rotation is found in equation 8, and $P_2(z)$, in equation 9.

$$z=s\epsilon^{-j\theta}; \quad s=z\epsilon^{j\theta} \quad (8)$$

and

$$P_2(z)=a_n\epsilon^{jn\theta}z^n+a_{n-1}\epsilon^{j(n-1)\theta}z^{n-1}+\ldots+a_1\epsilon^{j\theta}z+a_0 \quad (9)$$

The zeros of $P_2(z)$ are shown in Fig. 4 for the typical third-order polynomial. The coefficients of $P_2(z)$ are also complex, and the Hurwitz determinant cannot be used directly.

A polynomial in z containing real coefficients and the rotated zeros of $P(s)$ is obtained by multiplying the previous two polynomials in z:

$$F(z)=P_1(z)P_2(z) \quad (10)$$

The polynomial $F(z)$ contains all the zeros of $P_1(z)$ and $P_2(z)$. The locations of these zeros for the typical case considered in Figs. 3 and 4 are shown in Fig. 5. Due to the rotation of the original zeros of $P(s)$ in opposite directions, the zeros of $F(z)$ occur in complex conjugate pairs. Thus the coefficients of $F(z)$ are real.

Any pair of complex zeros of the original polynomial, $P(s)$, which lay outside the shaded sector in Fig. 2, are represented by a pair of complex zeros of $F(z)$ in the right half of the z-plane. Since $F(z)$ is of the order $2n$, the Hurwitz determinant of order $(2n-1)$ of the coefficients of $F(z)$ can then be used to detect zeros of $P(s)$ which exhibit less than a specified damping ratio. If the Hurwitz determinant for $F(z)$ is positive, then the zeros of $P(s)$ are located within the shaded region of Fig. 2. If the value is negative then at least two zeros of $P(s)$ are not sufficiently well damped.

Coefficients of F(z)

It has been shown that the coefficients of the new polynomial are real, but nothing yet has been shown about their values. The new coefficients are dependent upon the coefficients of $P(s)$ and also the angle of rotation θ. In general, the functional relationship in the fore-

---

Paper 57-742, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted November 14, 1956; made available for printing May 20, 1957.

THERON USHER, JR., is with Yale University, New Haven, Conn.

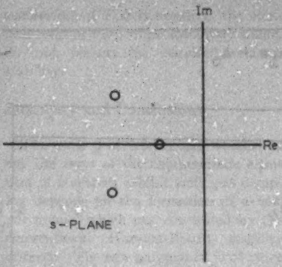


Fig. 1. Zeros of P(s)

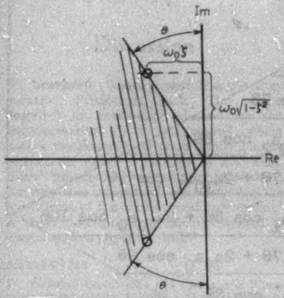


Fig. 2. Locus of constant ζ

going is found by performing the multiplication indicated in equation 10. The coefficients of $F(z)$ will be derived for the case where $P(s)$ is a third-order polynomial.

If

$$P(s)=a_3s^3+a_2s^2+a_1s+a_0 \tag{11}$$

then

$$P_1(z)=a_3\epsilon^{-j3\theta}z^3+a_2\epsilon^{-j2\theta}z^2+a_1\epsilon^{-j\theta}z+a_0 \tag{12}$$

and

$$P_2(z)=a_3\epsilon^{j3\theta}z^3+a_2\epsilon^{j2\theta}z^2+a_1\epsilon^{j\theta}z+a_0 \tag{13}$$

When equations 12 and 13 are multiplied the following are obtained:

$$\begin{aligned}F(z)=a_3^2z^6&+a_3a_2(\epsilon^{j\theta}+\epsilon^{-j\theta})z^5+\\&[a_2^2+a_3a_1(\epsilon^{j2\theta}+\epsilon^{-j2\theta})]z^4+\\&[a_2a_1(\epsilon^{j\theta}+\epsilon^{-j\theta})+a_3a_0(\epsilon^{j3\theta}+\epsilon^{-j3\theta})]z^3+\\&[a_1^2+a_2a_0(\epsilon^{j2\theta}+\epsilon^{-j2\theta})]z^2+\\&a_1a_0(\epsilon^{j\theta}+\epsilon^{-j\theta})z+a_0^2\end{aligned} \tag{14}$$

When the identity

$$\cos m\theta=\frac{\epsilon^{jm\theta}+\epsilon^{-jm\theta}}{2}$$

is recognized, then equation 14 becomes the following equation:

$$\begin{aligned}F(z)=a_3^2z^6&+2a_3a_2\cos\theta z^5+\\&(a_2^2+2a_3a_1\cos 2\theta)z^4+\\&(2a_2a_1\cos\theta+2a_3a_0\cos 3\theta)z^3+\\&(a_1^2+2a_2a_0\cos 2\theta)z^2+a_1a_0\cos\theta z+a_0^2\end{aligned} \tag{14A}$$

The coefficients of $F(z)$ for higher-order polynomials have also been derived. The results are listed in Table I for polynomials in s through the tenth order. Certain regularities in the subscripts are seen to occur, and thus the coefficients of $F(z)$ corresponding to any desired order of $P(s)$ may be formed. Consider the coefficient for z^{12}. The subscripts of the coefficients of $P(s)$ that are multiplied to form parts of the z^{12} term have a sum of 12. Both squared coefficients of $P(s)$ and also products of different coefficients of $P(s)$ appear in single coefficients of $F(z)$. The latter parts are always multiplied by twice the cosine of the difference of the coefficient indexes times θ, the angle of rotation (i.e., $a_7a_5\ 2\cos(7\text{-}5)\theta$).

Numerical Example

To illustrate the execution of the method, the polynomial in equation 15 will be examined for zeros which may have relative damping factors less than 0.5.

$$P(s)=s^3+5s^2+24s+20 \tag{15}$$

First, absolute stability is assured because all the coefficients are positive, and the Hurwitz determinant of the coefficients in equation 16 is positive:

$$D_{2(s)}=\begin{vmatrix}24 & 20\\ 1 & 5\end{vmatrix}=100>0 \tag{16}$$

For $\zeta=0.5$, $\theta=30°$, and
$\cos\theta=0.866$; $\cos 2\theta=0.500$; $\cos 3\theta=0$.

From Table I or equation 14(A) $F(z)$ is calculated (see equation 17 following).

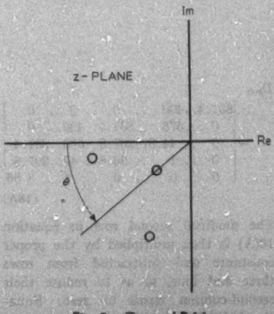


Fig. 3. Zeros of $P_1(z)$

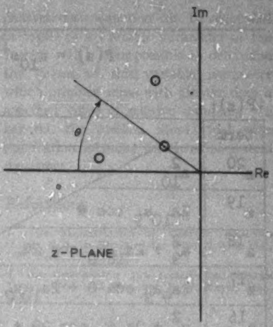


Fig. 4. Zeros of $P_2(z)$

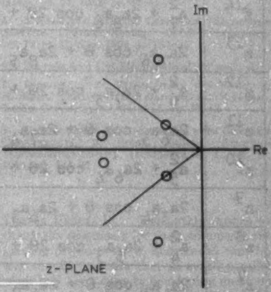


Fig. 5. Zeros of F(z)

$$\begin{aligned}F(z)=z^6&+2(1)(5)(0.866)z^5+\\&[5^2+2(1)(24)(0.50)]z^4+\\&[2(5)(24)(0.866)+0]z^3+\\&[24^2+2(5)(20)(0.50)]z^2+\\&2(24)(20)(0.866)z+20^2\end{aligned} \tag{17}$$

$$F(z)=z^6+8.66z^5+49z^4+207.8z^3+676z^2+831.4z+400$$

Next, $F(z)$ must be examined for zeros in the right half of the z-plane by evaluating the Hurwitz determinant of the coefficients of $F(z)$. The determinant is of the fifth order, and is found in equation 18.

$$D_{5(z)}=\begin{vmatrix}831.4 & ..400.. & 0 & .. & 0.. & 0\\ 207.8 & ..676.. & 831.4 & .. & 400.. & 0\\ 8.66.. & 49.. & 207.8 & .. & 676.. & 831.4\\ 0 & .. & 1.. & 8.66 & .. & 49..207.8\\ 0 & .. & 0.. & 0 & .. & 1..\ 8.66\end{vmatrix} \tag{18}$$

Since there are a great many zeros in this determinant, reduction to the triangular form is proposed as the quickest method of evaluation.

Table I. Coefficients of F(z)

$$P(s) = a_{10}s^{10} + a_9s^9 + a_8s^8 + \cdots + a_2s^2 + a_1s + a_0$$

F(z):

term	coefficient
z^{20}	a_{10}^2
z^{19}	$2a_{10}a_9 \cos\theta$
z^{18}	$a_9^2 + 2a_{10}a_8 \cos 2\theta$
z^{17}	$2a_9a_8 \cos\theta + 2a_{10}a_7 \cos 3\theta$
z^{16}	$a_8^2 + 2a_9a_7 \cos 2\theta + 2a_{10}a_6 \cos 4\theta$
z^{15}	$2a_8a_7 \cos\theta + 2a_9a_6 \cos 3\theta + 2a_{10}a_5 \cos 5\theta$
z^{14}	$a_7^2 + 2a_8a_6 \cos 2\theta + 2a_9a_5 \cos 4\theta + 2a_{10}a_4 \cos 6\theta$
z^{13}	$2a_7a_6 \cos\theta + 2a_8a_5 \cos 3\theta + 2a_9a_4 \cos 5\theta + 2a_{10}a_3 \cos 7\theta$
z^{12}	$a_6^2 + 2a_7a_5 \cos 2\theta + 2a_8a_4 \cos 4\theta + 2a_9a_3 \cos 6\theta + 2a_{10}a_2 \cos 8\theta$
z^{11}	$2a_6a_5 \cos\theta + 2a_7a_4 \cos 3\theta + 2a_8a_3 \cos 5\theta + 2a_9a_2 \cos 7\theta + 2a_{10}a_1 \cos 9\theta$
z^{10}	$a_5^2 + 2a_6a_4 \cos 2\theta + 2a_7a_3 \cos 4\theta + 2a_8a_2 \cos 6\theta + 2a_9a_1 \cos 8\theta + 2a_{10}a_0 \cos 10\theta$
z^{9}	$2a_5a_4 \cos\theta + 2a_6a_3 \cos 3\theta + 2a_7a_2 \cos 5\theta + 2a_8a_1 \cos 7\theta + 2a_9a_0 \cos 9\theta$
z^{8}	$a_4^2 + 2a_5a_3 \cos 2\theta + 2a_6a_2 \cos 4\theta + 2a_7a_1 \cos 6\theta + 2a_8a_0 \cos 8\theta$
z^{7}	$2a_4a_3 \cos\theta + 2a_5a_2 \cos 3\theta + 2a_6a_1 \cos 5\theta + 2a_7a_0 \cos 7\theta$
z^{6}	$a_3^2 + 2a_4a_2 \cos 2\theta + 2a_5a_1 \cos 4\theta + 2a_6a_0 \cos 6\theta$
z^{5}	$2a_3a_2 \cos\theta + 2a_4a_1 \cos 3\theta + 2a_5a_0 \cos 5\theta$
z^{4}	$a_2^2 + 2a_3a_1 \cos 2\theta + 2a_4a_0 \cos 4\theta$
z^{3}	$2a_2a_1 \cos\theta + 2a_3a_0 \cos 3\theta$
z^{2}	$a_1^2 + 2a_2a_0 \cos 2\theta$
z^{1}	$2a_1a_0 \cos\theta$
z^{0}	a_0^2

Added terms for higher order P(s)

If the first row of the determinant is multiplied by the proper constants, and then subtracted from the second and the third rows, the first column coefficients of the second and third rows are reduced to zero. The other terms in the second and third rows are modified by the nonzero terms in row 1. The value of the determinant is not modified by these operations and becomes equation (18A) as follows:

$$D_{s(z)} = \begin{vmatrix} 831.4 & 400 & 0 & 0 & 0 \\ 0 & 576 & 831.4 & 400 & 0 \\ 0 & 44.8 & 207.8 & 676 & 831.4 \\ 0 & 1 & 86.6 & 49 & 207.8 \\ 0 & 0 & 0 & 1 & 8.66 \end{vmatrix} \quad \text{(18A)}$$

The modified second row in equation 18(A) is then multiplied by the proper constants and subtracted from rows three and four, so as to reduce their second-column terms to zero. Equation 18(B) is the result.

$$D_{s(z)} = \begin{vmatrix} 831.4 & 400 & 0 & 0 & 0 \\ 0 & 576 & 831.4 & 400 & 0 \\ 0 & 0 & 143.1 & 644.8 & 831.4 \\ 0 & 0 & 85.2 & 48.3 & 207.8 \\ 0 & 0 & 0 & 1 & 8.66 \end{vmatrix} \quad \text{(18B)}$$

Two similar operations result in the desired triangular form:

$$D_{s(z)} = \begin{vmatrix} 831.4 & 400 & 0 & 0 & 0 \\ 0 & 576 & 831.4 & 400 & 0 \\ 0 & 0 & 143.1 & 644.8 & 831.4 \\ 0 & 0 & 0 & -335.7 & -287.2 \\ 0 & 0 & 0 & 0 & 7.77 \end{vmatrix} \quad \text{(18C)}$$

The value of the determinant is then the product of all terms along the primary diagonal.

$$D_{s(z)} = (831.4)(576)(143.1)(-335.7)(7.77) < 0 \quad \text{(18D)}$$

Since the Hurwitz determinant for the

coefficients of $F(z)$ is negative, the complex roots of $P(s)$, while absolutely stable, do not possess the required relative stability.

Summary and Conclusions

The method of determining whether or not the zeros of the characteristic equation of a system exhibit sufficient damping depends on the formation of a new polynomial which may be tested by the conventional Hurwitz-Routh stability criteria. The new polynomial is of twice the order of the original polynomial. The Hurwitz determinant for testing the new polynomial is fairly large, but the evaluation is simplified by reduction to triangular form.

The idea behind the method is thus rather simple, although the mathematical manipulations themselves tend to become cumbersome.

Immediate application of the method, as noted before, is found in the field of control system analysis. The relative stability of a proposed control system design can be examined by working with the characteristic equation of the closed-loop control system. The method may also be used directly in designing a control system. The coefficients of the characteristic equation can be expressed in terms of a variable parameter of the system. It is then possible to determine the values of this variable parameter which produce zeros with greater than a specified value of relative damping. In general, the method should find wide application in many problems of system dynamics.

References

1. The Mathematics of Circuit Analysis (book), E. A. Guillemin. John Wiley & Sons, Inc., New York, N. Y., 1949.

2. Servomechanism and Regulating System Design, Vol. I, H. Chestnut, R. W. Mayer. *Ibid.*, 1951.

◆

Discussion

W. Leonhard (Westinghouse Electric Corporation, Pittsburgh, Pa.): The author describes an interesting method for the transformation of a polynomial which makes the Hurwitz criterion applicable for testing the relative stability of a system. Lüthi proposed a very similar procedure, arriving also at general equations for the coefficients of the transformed polynomial.[1] A short description of Lüthi's method is contained in a paper by Görk.[2]

References

1. Damping Conditions for Regulator Equations of any Order, A. Lüthi. *Escher Wyss News*, Zurich, Switzerland, vols. 15, 16, 1942–43, pp. 90–95.

2. Stabilitätskriterien, E. Görk. *Archiv der Elektrischen Übertragung*, Berlin, Germany, no. 4, 1950, pp. 89–96.

N. H. Choksy (The Johns Hopkins University, Baltimore, Md.): The author's application is not new, since, during the past six to nine years, others have reported on it.[1-4]

If the function $P(s)$ of the author's equation 1 has real coefficients, it suffices to make one of the substitutions of his equations 6 and 8. This leads to a $P_1(Z)$ or a $P_2(Z)$ which has complex coefficients. Extensions of both the Hurwitz and the Routh criteria exist[7,8] for such polynomials, and in the former case they lead to the evaluation of determinants of no higher order than those needed in the author's procedure.

The author seems to imply that if the inequalities in his equation 2 are satisfied, the sufficient condition for stability reduces to his equation 3, i.e., the $(n-1)$th Hurwitz determinant should be positive. While this is true for $n=2$, 3, 4, for higher values of n, more of these determinants should be positive, e.g., for $n=5$, the sufficient conditions are that both D_2 and D_4 be positive.

References

1. Stabiliättskriterium Insbesondere von Regelkreisen bei Vorgeschriebenes Stabilitätsgute, A. Leonhard. *Archiv der Elektrischen Übertragung*, Berlin, Germany, vol. 39, 1948, pp. 100–07.

2. A Generalization of Nyquist's Stability Criterion, A. Vazsonyi. *Journal of Applied Physics*, New York, N. Y., vol. 20, 1949, pp. 863–67.

3. Relative Damping as Criterion for Stability and as an Aid in Finding the Roots of a Hurwitz Polynomial, A. Leonhard. "Automatic and Manual Control" (book), A. Tustin, editor. Butterworths Scientific Publications, London, England, 1952, pp. 25–35.

4. Koeffizientenbedingungen zur Kontrolle des Dämpfungsgrades bei Ausgleichvorgängen (Verallgemeinerte Hurwitzbedingungen), O. Schmutz. *Ingenieurarchiv*, Berlin, Germany, vol. 21, 1953, pp. 33–41.

5. A Relative Damping Criterion for Linear Systems, J. Frank Koenig. *AIEE Transactions*, vol. 72, pt. II, Nov. 1953, pp. 291–95.

6. See reference 1 of Mr. Leonhard's discussion.

7. Bemerkung zu einem Staze von Hurwitz, H. Billars. *Zeitschrift für Mathematik and Mechanik*, Berlin, Germany, vol. 24, 1944, pp. 77–82.

8. On the Zeros of Polynomials with Complex Coefficients, E. Frank. *Bulletin*, American Mathematical Society, Providence, R. I., vol. 52, 1946, pp. 144–57.

Theron Usher, Jr.: The author wishes to thank the discussers for their comments regarding previous work, and to say that he was simply unaware that so much parallel work had been done in developing methods for determining the relative stability of the zeros of polynomials. The graphic methods of Leonhard and Vazsonyi, and the analytical method of Frank and Koenig are particularly interesting.

The author wishes to thank Mr. Choksy for clearing up the misconception concerning the number of Hurwitz determinants that must be positive as a sufficient condition for stability.

The same comment should be applicable to the Hurwitz determinant that contains the coefficients of $F(z)$.

Magnetically Regulated Reference Power Supply

ROBERT A. RABER
NONMEMBER AIEE

THE NEED for a highly accurate and reliable d-c voltage reference as part of a Navy Gun Fire-Control System resulted in development of the reference supply described. The customer's requirements imposed severe size and reliability restrictions. The reliability requirements resulted in the choice of magnetic amplifiers for instrument servos of the system. Upon evidence of the feasibility of their use, they were also chosen as logical components for the reference power supply.

A ferroresonant regulator was used. This device produces an output voltage that is nearly a square wave and thus contains little of the lower harmonics of the line voltage after rectification. This keeps the size of the output filter to a minimum. The regulator is relatively insensitive to changes in line voltage and thus simplifies the job to be done by the regulating loop. The only reference that could promise sufficient accuracy combined with long reliability was the silicon-diode reference. The one chosen has a specified accuracy of 0.001% per degree centigrade (C) from −55 C to +100 C. The proper combination of these components, with the application of basic servo techniques, resulted in a compact, reliable, and accurate reference supply.

Performance

The reference power supply (Fig. 1) provides an output voltage of 250 volts d-c, accurate to 0.1% for the following conditions: line voltage 115 volt ±10%, 400 cycles per second ±5%; ambient temperature −25 C to +55 C. It has been designed with an expected life of 5 years, with 1,000 hours of operation per year. Varying the output load from 15 to 150 milliamperes (ma) results in an output variation of 50 millivolts. Output ripple is 15 millivolts root mean square. Data on transient response to steps of line voltage and load current are presented in Fig. 2. The production version of the reference supply has been designed to occupy a volume 7 by 7 by 4 inches. Estimated weight is 10.5 pounds using acceptable Navy construction.

Operation

The regulator is of the ferroresonant type. It consists of a nonlinear reactor in parallel with a capacitor. The voltage developed across this combination is essentially a square wave for properly chosen components. The theory of operation is explained in the appendix.

A magnetic amplifier in series with the nonlinear, or saturating reactor of the ferroresonant regulator, controls the amplitude of the square wave of voltage. The saturating reactor is called the main stabilizer. The square-wave voltage is rectified in a conventional bridge rectifier and then filtered by a 2-stage inductance-capacitance filter. A resistance bridge compares a portion of the output voltage with the silicon-diode reference voltage to obtain an error signal.

The error is amplified in a full wave, push-pull, d-c output, magnetic preamplifier stage. Positive feedback on this stage is used as a loop gain control. Excitation for this stage is taken from a winding on the saturating reactor of the ferroresonant regulator. Using square-wave excitation voltage results in greater linearity of the gain characteristic of the preamplifier.

The amplified error signal is used to control the magnetic amplifier stage which is in series with the saturating reactor (main stabilizer) of the ferroresonant circuit. Thus, the error signal controls the amplitude of the voltage from the ferroresonant regulator. This series magnetic amplifier has been referred to as a "booster amplifier" by previous authors. It increases, or boosts, the level of the output voltage above that which would be available from the ferroresonant regulator if only the main stabilizer were used. The booster, a single-ended a-c stage, is biased to operate on the linear portion of its gain characteristic. The bias current is taken from the output of the power supply. It is adjusted to set the regulator output at the desired nominal setting of the power supply output voltage. The precise adjustment of the output voltage is made in the resistance bridge which develops the error signal.

Because of the polarity of the bias required there is a positive feedback effect on the booster stage. This positive feedback increases the effective gain of the booster stage. Stable operation is obtained by use of an *R-C* network in the bias circuit. This attenuates the positive feedback at high rates of change of output. To get improved transient response, a rate feedback signal is inserted at the preamplifier. This is an *R-C* circuit which results in a feedback current whenever changes occur in output voltage.

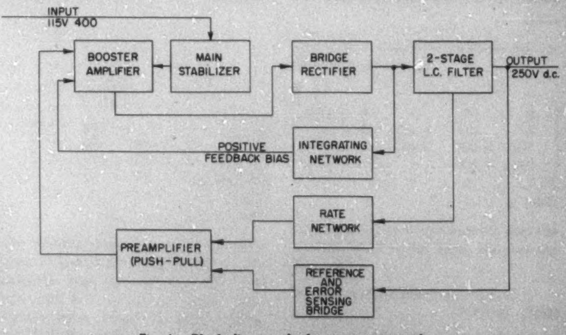


Fig. 1. Block diagram of reference power supply

Paper 57-770, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted October 25, 1956; made available for printing May 21, 1957.

ROBERT A. RABER is with the General Electric Company, Schenectady, N. Y.

The author wishes to acknowledge the earlier work of Dr. E. Manteuffel and R. O. McCary of the General Electric Advanced Electronics Center, Cornell University, which was of immeasurable help to the successful completion of this development. Work on this project was done in conjunction with C. C. Thompson of General Electric Company, Schenectady, N. Y.

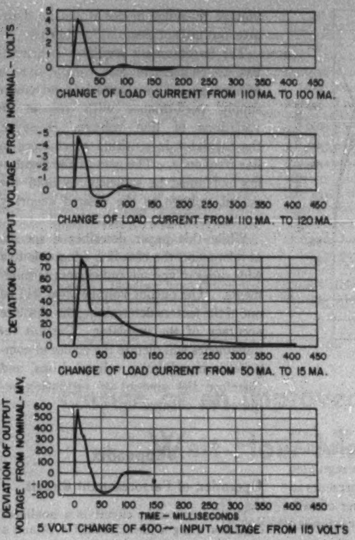


Fig. 2 (left). Transient response for step changes of load current and line voltage

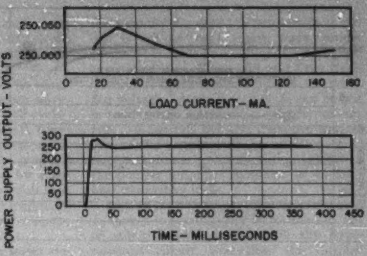


Fig. 4 (above). Power supply output voltage versus load current and versus time (starting transient)

Addition of the rate signal allows stable operation with much higher amplifier gain than would otherwise be possible.

Experimental Results

Fig. 3 presents data showing the variation of output voltage for line voltage, frequency, and temperature variations. These data were taken without subjecting the diode reference, and the error sensing resistors R_9 and R_{10}, to the temperature variations.

The variations measured under these conditions reached a maximum of ±0.022%. The reference diode is manufactured to a specified accuracy of 0.001% per C which contributes an additional error of ±0.04% for a total of ±0.062%. This leaves some margin for errors which might be contributed from R_9 and R_{10}. These two resistors must be specially constructed. R_9 is 25,000 ohms, R_{10} is 750 ohms. They must have the same temperature coefficient of resistance, and must be properly rated so that selfheating effects are equal. They need not be very accurate in absolute value, however, as a potentiometer, P_4, makes the final setting. The purpose of these resistors is to divide the output voltage down to a level where a comparison can be made with the reference diode voltage. The network R_9, P_4, and R_{10} is being manufactured to these requirements by a commercial supplier.

Fig. 2 illustrates the transient response of the power supply to step changes of load current and line voltage. Data are presented as deviation of output voltage from nominal versus time in milliseconds. The output voltage returns to less than 0.5% (1.25 volts) within 30 milliseconds for a 10-ma step change of load current. Output voltage returns to 0.25 volt, or 0.1% in 100 milliseconds or less. A step change of 5 volts in the 400 cycles per second supply voltage results in a transient in the output voltage which never exceeds 600 millivolts, or 0.24%, and returns to 0.1% (250 millivolts) in less than 25 milliseconds.

Fig. 4 shows the change in output voltage as load current is changed from 15 to 150 ma. The maximum change that occurs is 50 millivolts. No change in output voltage could be detected over the range from 70 to 130 ma. The second illustration of Fig. 4 records the starting transient, or output voltage versus time (milliseconds) when the line switch to the supply is closed. Note that output voltage is within the regulating accuracy within a fraction of a second. The absence of warm-up time is a distinct advantage of the magnetic regulated supply. The loop gain is approximately 350 volts per volt. The ferroresonant circuit output is directly proportional to line frequency. Thus the loop gain is a measure of the power supply's ability to hold accuracy when line frequency variations occur.

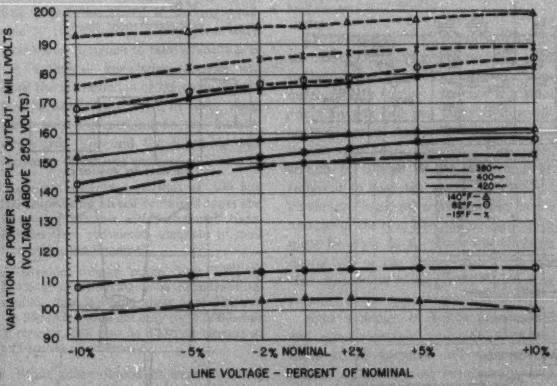


Fig. 3. Variation of power supply output voltage for changes in line voltage, line frequency, and ambient temperature

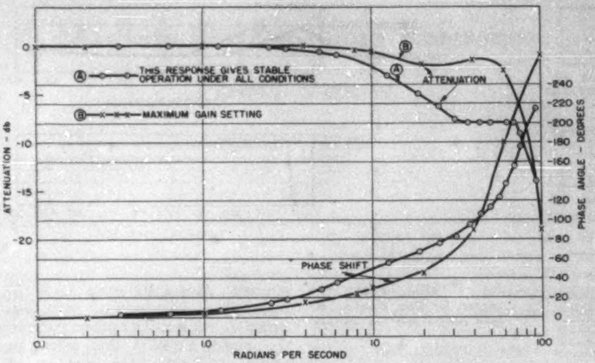


Fig. 5. Closed-loop frequency response of reference power supply

The closed loop frequency response is plotted in Fig. 5. Two conditions are illustrated. Curve *A* represents a gain setting which will yield stable operation over all possible combinations of line and ambient conditions. Curve *B*, for a higher gain setting, would be a more desirable response. Gain and response changes which occur for combined high frequency and low temperature will result in instability, however, with the gain setting of curve *B*.

Conclusions

The dependability and freedom from maintenance that results from the use of magnetic amplifier components have made them eminently suited for application to military systems. The regulated d-c voltage supply described represents an application of magnetic amplifier circuits to provide a reliable, high-power reference supply.

The current capabilities of magnetically regulated supplies makes them superior in size to electronic supplies having conventional commercial reliability. The data demonstrate the ability of this type of supply to give excellent accuracy with good transient performance. The high accuracy and stability of the new silicon reference diodes make them ideal references for this type of supply. Higher voltage reference diodes would be desirable, but are not essential for this use.

The ferroresonant regulator is insensitive to changes in line voltage over a wide range. This results in good power supply performance with respect to variations of line voltage. However, the regulator is sensitive to changes in line frequency. The gain of the regulating loop must be sufficient to hold the output voltage to the desired accuracy for line frequency variations.

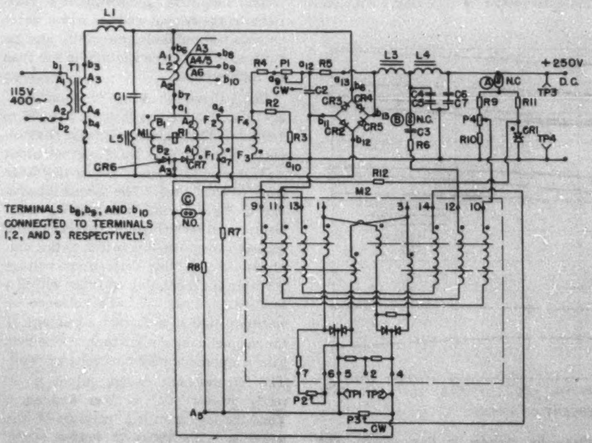
While this paper describes a specific development, the circuit is adaptable to a wide range of voltage and current requirements. The output voltage level will be one of the factors determining the absolute accuracy of the regulator. Current requirements will dictate the choice of components to minimize voltage drops, and therefore the amount of regulation required for changes in load.

Appendix

Operation of Ferroresonant Circuit

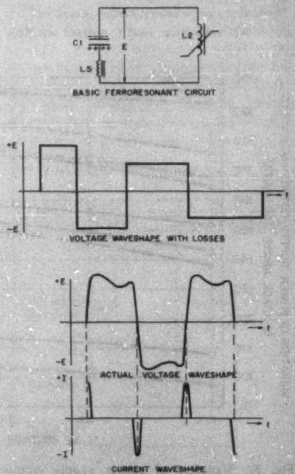
The ferroresonant circuit is a nonlinear oscillator using a linear capacitor C_1 and a saturating reactor L_2 with a square loop core material; see Fig. 6. Linear reactor L_1 keeps the source impedance high, as seen from the ferroresonant circuit. The booster amplifier M_1 can be neglected for the moment.

Referring to Fig. 7, let us consider C_1 and L_2 of Fig. 6 as if they were isolated from the rest of the circuit. If C_1 is charged at voltage $+E$ it will try to discharge through L_2, but as long as L_2 is not saturated, current is



L1
T1
115V 400~
C1
L5
M1
L2
CR6
CR7
R1
R2
R3
R4
P1
R5
CW
C2
CR1
CR2
CR3
CR4
CR5
L3
L4
C3
C4
C5
C6
C7
R6
R9
R10
R11
P4
N.C.
+250V D.C.
TP3
TP4
R12
M2
R7
R8
N.O.
P2
P3
TP1 TP2
A5
TERMINALS b_8, b_9, AND b_{10} CONNECTED TO TERMINALS 1, 2, AND 3 RESPECTIVELY.

Fig. 6. Electrical schematic diagram of reference power supply



C1
E
L2
L5
BASIC FERRORESONANT CIRCUIT
+E
-E
t
VOLTAGE WAVESHAPE WITH LOSSES
ACTUAL VOLTAGE WAVESHAPE
+I
-I
CURRENT WAVESHAPE

Fig. 7. Ferroresonant regulator

limited to the magnetizing current of L_2, which is quite small. This assumes that the resonant frequency of C_1 and unsaturated L_2 is very low. As a result flux in L_2 changes at a constant rate which will just sustain the voltage E across L_2. When L_2 reaches saturation it changes to a very low impedance (saturated inductance and winding resistance). This allows a large pulse of current to flow and reverse charge on C_1 to $-E$. The resonant frequency is now that of C_1 and the saturated inductance of L_2, a relatively high frequency. Voltage of negative polarity now appears across L_2 and magnetizing current flows in the opposite direction. Voltage of $-E$ will exist across L_2 until it reaches negative saturation and allows C_1 to discharge again. Thus we have completed a cycle during which we had voltage of constant amplitude E, first plus and then minus. This change occurs in a very short time. We have therefore described a square wave of voltage across L_2.

In a practical case there would be some losses. The amplitude of the square wave would therefore decrease on each half cycle. Since the total core flux is the same, however, the lower voltage would result in a slower rate of change of flux. The product of volts and time (volt-time area) for each half cycle will be the same. The result is a square wave, each half cycle of which is somewhat lower in amplitude and proportionately longer in period.

In the power supply circuit, the losses are replaced from the line each half cycle. Linear reactor L_1 prevents short-circuiting the line through L_2 during saturation. While the output voltage of the ferroresonant circuit is not an ideal square wave, proper choice of components will result in a good approximation. There will be a very small amount of fundamental ripple in the rectified voltage. A small reactor L_3 can be put in series with C_1 to limit the discharge current. While this detracts from the squareness of the output voltage, addition of L_3 may be necessary to stay within ratings of the booster amplifier rectifiers, or to limit heat dissipation in L_2.

Adding the booster amplifier in series with L_2 makes it possible to control the amount of volt-time area available in the ferroresonant circuit, and thus the output-voltage (square-wave) amplitude. This controls the level of the power-supply output voltage.

Design of an Improved 2-Stator Watt-Hour Meter

D. M. SMITH
NONMEMBER AIEE

Synopsis: An improved design of 2-stator polyphase watt-hour meter is described in which low rotor speed is utilized to give improved performance over extended load and voltage ranges. Fundamental performance equations are derived to show how this improved performance is attained.

When it became evident in recent years that existing designs of polyphase watt-hour meters would shortly be made obsolete by changing load requirements and technological advances in the art of metering, a development program was undertaken aimed at making available to the industry, a polyphase watt-hour meter designed to take advantage of the technical knowledge accumulated since the introduction of the basic designs now in use. As a result of early work, effort was concentrated on the 2-stator single-disk type and the following performance objectives were set up:

Maximum current capacities of 100 and 200 amperes for all self-contained types and 10 amperes for use with instrument transformers were considered adequate to meet foreseeable requirements.

A useable voltage range from 50% to 130% of rated voltage was considered desirable, since it would allow application of the 240-volt rating not only on 277/480-volt Y circuits, but also on 120-volt circuits at half-torque, if so desired.

While other objectives were achieved, such as reduced maintenance cost and improved stability of calibration, they are outside the scope of this paper.

Extended Load Range

The extension of the useable overload range of a nominally rated 15-ampere stator from the previous 50-ampere limit to 100-ampere capacity involves two types of design problems; namely,

1. Thermal, i.e., keeping the maximum hot-spot temperature within the long-life operating range of the insulating materials used.

2. Accuracy, i.e., maintaining close to 100% accurate registration throughout the load range.

The thermal problem requires first that the current circuit ampere-turns be kept to the minimum commensurate with adequate driving torque; yet, not less than one turn per coil per pole. For a 15-ampere 3-wire stator, 60 ampere-turns works out well. Second, the cross section of copper in the coils must be sufficient to keep the watts generated at maximum load at a reasonable level; and third, the lead size and configuration must be such as to allow the effective dissipation of the heat through conduction and convection without exceeding permissible hot-spot temperatures.

Since a design of 3-wire current coil and core assembly giving 60 ampere-turns at 15 amperes and meeting the second of the foregoing requirements was available from the extended range single-phase meter, it was possible to design current leads for attachment to this assembly such that even with two stators operating at 100 amperes in a single case, the hot-spot temperature rise does not exceed the permissible limit for the coil insulation used.

The use of the 3-wire single-phase meter coils served to fill the need not only for the 15-ampere 2-wire stators, by connecting the two coils in series, but also for the 15-ampere 3-wire stators used in 4-wire Y and delta meters, by using the coils separately, and for the 30-ampere 2-wire stators by connecting the two coils in parallel.

The 30-ampere 3-wire stator is an entirely different problem. It appears at present that the logical solution to the difficulties encountered with orthodox designs of 50- or 30-ampere 3-wire current coils is to adopt the construction described by Landry and Scamman.[1]

The solution to the problem of accuracy is more involved. Since a 2-stator meter on a polyphase circuit may operate with only one current circuit loaded, as in the case of a 4-wire delta circuit with load only on the lighting phase, it is necessary to consider the performance under these conditions as well as with both stators loaded. To evaluate the effect of this circumstance on performance, consider the fundamental factors which determine the shape of the overload accuracy curve.

Let

T = driving torque per stator
T_i = retarding torque per stator due to current flux
T_e = retarding torque per stator due to potential flux

Paper 57-792, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 27, 1957; made available for printing June 11, 1957.

D. M. Smith is with the General Electric Company, Somersworth, N. H.

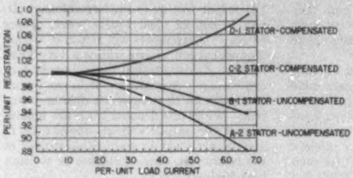


Fig. 1. Calculated inherent overload curves for $16^2/_3$-rpm meter, 1.0 power factor

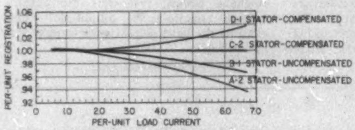


Fig. 2. Calculated inherent overload curves for $8^1/_3$-rpm meter, 1.0 power factor

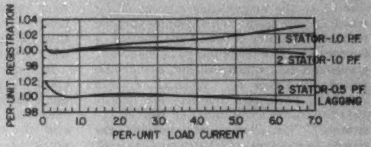


Fig. 3. Compensated load curves for $8^1/_3$-rpm meter

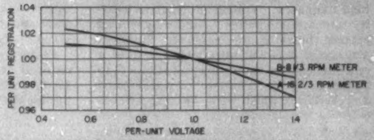


Fig. 4. Calculated inherent voltage curves at rated current, 1.0 power factor

A—$16^2/_3$-rpm meter
B—$8^1/_3$-rpm meter

T_m = retarding torque due to retarding magnet
S = rotor speed
(The foregoing are all at rated load and voltage.)
X = per-unit load current, i.e., at rated load $X=1.0$
T_x = driving torque per stator at per-unit load X
S_x = rotor speed at per-unit load X

Then, for a 2-stator meter with both stators operating at rated load and voltage and unity power factor (neglecting friction torque, which is insignificant at other than very light loads)

$$2T=2T_i+2T_e+T_m \qquad (1)$$

Since the potential flux is constant at constant voltage, the retarding torque due to potential flux can be considered the same as that of the retarding magnet, i.e., the total retarding torque due to these two components will be proportional to rotor speed only, at rated voltage.

Hence

$$2T=2T_i+KS \qquad (2)$$

and

$$K=\frac{2(T-T_i)}{S} \qquad (3)$$

where K is a constant for a given design at rated voltage.

For any load X the following applies:

1. Driving torque T_x is assumed proportional to load current, i.e., the current-flux magnetic circuit is assumed linear and the meter uncompensated.
2. Retarding torque due to current flux is proportional to the square of the load current and is directly proportional to rotor speed.
3. Retarding torque due to potential flux and retarding magnet is directly proportional to rotor speed only at constant voltage.

Therefore, at any per-unit load X_i

$$2T_x=2XT=2X^2\frac{(S_x)}{S}T_i+KS_x \qquad (4)$$

Solving for S_x and simplifying gives the following equation

$$S_x=\frac{2XT}{\frac{2X^2T_i}{S}+K}=\frac{2XT}{\frac{2X^2T_i}{S}+\frac{2(T-T_i)}{S}}=\frac{X}{1+(X^2-1)T_i/T}\times S \qquad (5)$$

and

$$\frac{S_x}{S}=\frac{X}{1+(X^2-1)T_i/T} \qquad (6)$$

Since the correct speed at any load X would be XS, the per-unit registration at load X is given by

$$\text{per-unit registration}=\frac{\text{actual speed}}{\text{correct speed}}=\frac{S_x}{XS}=\frac{1}{1+(X^2-1)T_i/T} \qquad (7)$$

To determine the actual shape of this overload curve, a value of $T_i/T=0.00313$ is assumed, which gives 12% droop at $666^2/_3$% rated load, a realistic value for an uncompensated meter having a rotor speed of $16^2/_3$ rpm at rated load. This gives curve A in Fig. 1.

Now consider the effect when only one stator is loaded. The driving torque is halved, as is the current flux, but since the rotor speed also decreases by about one half, the retarding torque due to the current flux is reduced to much less than one half as given in the following equation:

$$T_x=XT=X^2\frac{S_x}{S}T_i+KS_x \qquad (8)$$

Solving for S_x/S as before

$$\frac{S_x}{S}=\frac{X}{2+(X^2-2)T_i/T} \qquad (9)$$

Since with one stator loaded the correct speed would be $XS/2$, thus

$$\text{per-unit registration}=\frac{S_x}{\frac{XS}{2}}=\frac{S_x}{S}\times\frac{2}{X}=\frac{1}{1+\frac{(X^2-2)}{2}T_i/T} \qquad (10)$$

Using the same value for $T_i/T=0.00313$ as before, this equation gives curve B in Fig. 1. Note that curve B droops only about 6.3% at $666^2/_3$% load, giving a spread of 5.7% between 2-stator and single-stator operation.

Now overload compensation is applied to determine the effect on these fundamental equations. It is assumed that the compensation causes a straight-line overload characteristic for 2-stator operation, a condition which simplifies the mathematics and can be approached in practice if desired by proper compensation.

Since overload compensation causes the current flux to increase at a greater rate than the load current, it may no longer be said that the current-flux retarding torque varies as X^2, but for X must be substituted the actual ratio of current flux at load X to that at rated load, as determined by the driving torque itself, i.e.,

$$\text{current-flux retarding torque at load } X=\left(\frac{T_x}{T}\right)^2\left(\frac{S_x}{S}\right)T_i \text{ per stator} \qquad (11)$$

Also XT may not be substituted for

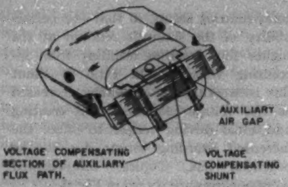

AUXILIARY AIR GAP
VOLTAGE COMPENSATING SECTION OF AUXILIARY FLUX PATH.
VOLTAGE COMPENSATING SHUNT

Fig. 5. Potential coil and core assembly showing arrangement of voltage-compensating shunt

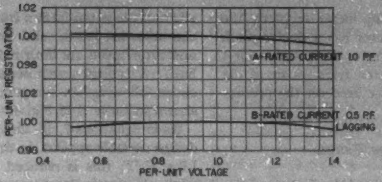


Fig. 6. Compensated voltage curves for $8^1/_3$-rpm meter

A—Rated current 1.0 power factor
B—Rated current 0.5 power factor lagging

T_x, since T_x is not proportional to X. The fundamental equation for 2-stator operation is then

$$2T_x=2\left(\frac{T_x}{T}\right)^2\left(\frac{S_x}{S}\right)T_i+KS_x \tag{12}$$

However, it was stated that compensation is to be perfect, i.e.,

$$S_x=XS \tag{13}$$

Substituting this for S_x in equation 12

$$2T_x=2\left(\frac{T_x}{T}\right)^2 XT_i+KXS \tag{14}$$

Substituting the value of KS from equation 2 and rearranging gives

$$X\frac{T_i}{T}\left(\frac{T_x}{T}\right)^2-\left(\frac{T_x}{T}\right)+X(1-T_i/T)=0 \tag{15}$$

Solving equation 15 for T_x/T

$$\frac{T_x}{T}=\frac{1\pm\sqrt{1-4X^2(T_i/T-T_i^2/T^2)}}{2XT_i/T} \tag{16}$$

Although this looks formidable, it is in terms of X, and T_i/T only and can readily be tabulated for values of T_x/T for later use. (By trial, only the minus root is valid.)

Now for one stator, with this same overload compensation, T_x/T remains the same as when both stators are loaded, since it is not sensitive to rotor speed but only to load X. Then, for one stator compensate

$$T_x=\left(\frac{T_x}{T}\right)^2\left(\frac{S_x}{S}\right)T_i+KS_x \tag{17}$$

Solving for S_x/S as before

$$\frac{S_x}{S}=\frac{T_x/T}{2+\left(\frac{T_x^2}{T^2}-2\right)T_i/T} \tag{18}$$

and

per-unit registration

$$=\frac{S_x}{S}\times\frac{2}{X}=\frac{T_x/T}{X\left[1+\left(\frac{T_x^2}{2T^2}-1\right)\frac{T_i}{T}\right]} \tag{19}$$

Substituting the values of T_x/T obtained from equation 16, using the same value of $T_i/T=0.00313$ as previously, results in curve D in Fig. 1. It may be seen that this curve has a rising characteristic such that at $666^2/_3$% load the plus error is 9%, thus also giving a spread of 9% between curves C and D for 2-stator and single-stator loading respectively.

It is now clear that the inherent spread in overload curves between 2-stator and single-stator loading is not reduced by compensating the meter; it is, in fact, increased rather markedly at the higher loads due to the over-all increase in both current flux and rotor speed resulting from compensation. Not serious in earlier limited range meters, this effect becomes a major problem in extended range designs which may operate at rather high loads even under single-stator conditions.

Since the only design variable in these fundamental performance equations is the quantity T_i/T, obviously the only way to reduce the spread between the 2-stator and single-stator load curves is to reduce T_i/T. While this factor can be reduced by changes in the geometry of the stator and rotor to make the current flux more effective in driving than in retarding, thus decreasing T_i/T, any substantial gain in this direction would require a rearrangement of components which would seriously compromise the mechanical design of the meter.

However, the desired objective was accomplished by increasing the strength of the retarding magnet so as to reduce the rotor speed at rated load, since T_i is reduced directly as the speed, while T remains constant. The reduction in rotor speed at rated load from $16^2/_3$ to $8^1/_3$ rpm cuts the factor T_i/T in half, and the use of equations 7, 10, and 19 then results in the curves shown in Fig. 2. Note that the spread between single-stator and 2-stator operation is less than one half that shown in Fig. 1 for $16^2/_3$-rpm rated speed, particularly at the higher loads.

As a final compromise, the meter was compensated so that the 2-stator load curve droops slightly at $666^2/_3$% load, thus reducing the rise of the single-stator curve to a more reasonable value as shown in the curves in Fig. 3. An additional benefit of the reduction in rated speed is that the total amount of overload compensation required is cut in half, thus reducing the effect on the load curve of normal manufacturing variations between meters.

To obtain the maximum interchange-

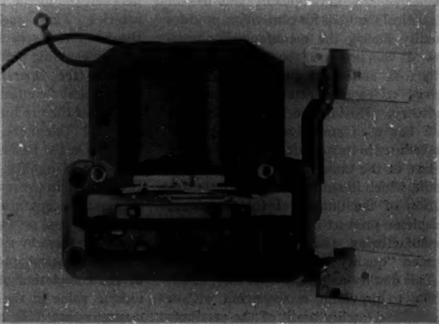


Fig. 7. Improved stator with extended load and voltage range

ability with existing meter registers, the new half-speed meter has a double-pitch drive worm in the rotor shaft so that for a given rating of meter, registers are of the same ratio as those used on the present $16^2/_3$-rpm meters.

Extended Voltage Range

A similar analysis of the factors affecting the inherent voltage curve results in the following relation for unity power factor at rated load without voltage compensation.

$$\text{per-unit registration} = \frac{1}{1+(X^2-1)T_e/T} \qquad (20)$$

where X = per-unit voltage, i.e., $X = 1.0$ at rated voltage.

The similarity to equation 7 will be noted, and it is logical since variation in either current flux or potential flux has a similar effect on retarding torques due to these fluxes, and hence on the performance characteristics. This relation applies also, approximately, to the condition of single-stator loading, as the retarding effect of the potential flux is still present in the unloaded stator, and driving torque and rotor speed are both reduced by one half. A typical uncompensated voltage curve following the foregoing relation is shown in curve *A* in Fig. 4.

As in equations 7, 10, and 19, only one design variable appears, in this case T_e/T. Like T_i/T, this is reduced by one half when the nominal rotor speed is cut in half by increasing the strength of the retarding magnet, and the result is shown by curve *B* in Fig. 4 for a nominal rotor speed of $8^1/_3$ rpm.

The reduction in inherent voltage error makes the meter easier to compensate and improves the uniformity of characteristics among meters. To cover the required useable range of 50% to 130% of rated voltage in this case, it was found necessary to add a saturating shunt across the air gaps in the auxiliary potential flux circuit, in addition to operating the iron in the auxiliary flux circuit itself above its maximum permeability.

The arrangement of the voltage-compensating shunt is shown in Fig. 5. Its action is similar to that of the usual load-compensating shunt, in that the permeability of the shunt is reduced at the higher flux densities resulting from increasing voltage. The ratio of shunt flux to working flux is thereby decreased as the voltage increases, thus creating additional driving torque to offset the increased retarding torque due to the potential flux. Typical compensated voltage curves are shown in Fig. 6.

Summary

An improved design of 2-stator single-disk watt-hour meter has been achieved (see Fig. 7), which has 100- and 200-ampere capacity in the 15-ampere and 30-ampere nominal ratings respectively. It is shown that accuracy under unbalanced loads is substantially improved, and uniformity of characteristics enhanced, by reducing the rotor speed at nominal rated load to one half of that now in common practice.

Reference

1. An Improved Single-Phase Watt-Hour Meter, J. G. Landry, J. F. Scamman. *AIEE Transactions*, vol. 76, pt. I, Sept. 1957, pp. 438–39.

Quality Assurance—An Essential Part of the Western Electric Manufacturing Process

I. O. KARRAKER
NONMEMBER AIEE

THE Bell Telephone Laboratories and the Western Electric Company (WE) pioneered the development and use of statistical methods for controlling product quality during the manufacturing process. It was not until the war years that this tool, statistical quality control, received widespread attention throughout industry. During the many years that WE has used statistical quality-control procedures in its manufacturing processes, it has, at the same time, been making an audit, which in reality is a statistical evaluation of the quality of the completed telephone product being shipped from its manufacturing plants. This audit is termed "quality assurance" in WE.

This quality-assurance approach, in addition to providing management with a continuing quality audit of the products it is shipping, further permits WE to furnish quality data on the various products shipped to the Bell Telephone Laboratories. These quality data are used by the laboratories in forming a picture of quality enabling them to provide the customers (the Operating Bell Telephone Companies) "continuing evidence of the quality of the products being delivered to the Bell Telephone Companies by the WE." On the basis of the receipt of this evidence of quality, the Bell Telephone Companies, in the main, find it possible to eliminate acceptance inspections and the associated cost on the products being delivered to them by the WE.

The statistical quality-assurance audit at the completion of manufacture has a unique value in the communications industry where WE, to a large extent, is both the maker and the installer of the product which make possible the vast Bell System network of local telephone switching equipment and long distance lines, repeaters, and relays. Instead of manufacturing being the completion of the company's job, it is but a single stage in the over-all process of furnishing communications' services to the nation. From this viewpoint then, quality assurance, prior to shipment of product, becomes part of a much larger pattern, the Bell System of communications' facilities. The statistical method by which this is accomplished has been developed through experience and study, and has been an important factor in achieving balance between quality of product and cost of manufacture. The object of this paper is to explain how this type of statistical control works, and how it is administered.

Statistical Approach

In considering the quality of a variety of complex telephone products on a sta-

---

Paper 57-808, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted October 17, 1956; made available for printing May 23, 1957.

I. O. Karraker is with the Western Electric Company, Omaha, Nebr.

tistical basis, it is necessary to establish certain common denominators both for determining a sample and for evaluating the results of the inspection performed upon that sample. Since there are so many kinds of items, and each type must be sampled separately, this means that an individual unit of sample may vary in size from a small and simple fuse to a large and complicated $11^1/_2$-foot crossbar central-office frame. Once a sample of a given item is chosen, the general purpose is the same; to determine whether the product conforms to the engineering requirements and standards of workmanship which have been established for it. The engineering requirements are established by the Bell Telephone Laboratories' development engineers, and the degree to which product must conform to these requirements is determined by the laboratories' quality-assurance organization, that acts as the customer's representative on all matters relating to quality.

As the aim is to keep a constant record of the quality as equipment passes from factory to installer, results of this engineering inspection are analyzed once a month, although samples for each product must be chosen throughout the month in order that they may be random and representative of the output. Size of the actual monthly sample depends upon output and may be different for various classes of product. In general, the larger the production, the smaller the percentage inspected in the quality-assurance sample. For small outputs the sample is normally 10%, and for very large outputs of highly repetitive items a maximum sample of possibly 500 has been set up.

With long experience then, a definite method of determining a monthly sample for all the various products has been evolved, and the aim, in all cases, is to see how well the factory is meeting the design intent of all manufactured output as it goes out for installation in the telephone plant. The information thus obtained enables management to judge the manufacturing portion of the Bell System from the standpoint of quality.

Classification of Defects

When a sample is selected, it is thoroughly inspected by the quality-assurance personnel. In this process, reference is made to drawings, specifications for the particular order, and any other information affecting the requirements. Checks are made of adjustments, defects in assembly, required markings, workmanship, damage, and any specified electrical items. If the particular sample includes wiring and connecting, a careful inspection is made of soldered connections and the associated wiring, including such matters as gage of wire and the correctness of wiring dress. Generally, an operation test is made which simulates as nearly as possible the conditions under which the product will function in service.

Defects discovered are of varying degrees of seriousness and the seriousness is measured by demerits used as a common denominator. In order to simplify evaluation of defects and to minimize the effect of personal judgment, a four-fold classification has been developed which is fundamental to all WE quality-assurance ratings. The classification is as follows.

Class A—Very Serious, Demerit Value 100

1. Will surely cause an operating failure of the unit in service which cannot be readily corrected in the field, e.g., open relay winding.

2. Will surely cause intermittent operating trouble difficult to locate in the field, e.g., loose soldered connection.

3. Will render unit totally unfit for service e.g., test of repeater shows no gain.

4. Liable to cause personal injury or property damage—under normal conditions of use, e.g., exposed part has sharp edges.

Class B—Serious, Demerit Value 50

1. Will probably cause an operating failure of the unit in service which cannot be readily corrected in the field, e.g., protective finish missing from coaxial plug.

2. Will surely cause an operating failure of the unit in service which can be readily corrected in the field, e.g., relay contact does not make.

3. Will surely cause trouble of a nature less serious than an operating failure, such as substandard performance, e.g., amplifier regulation more than 10% outside limits.

4. Will surely involve increased maintenance or decreased life, e.g., single contact disk missing or contact portection network missing.

5. Defects of appearance or finish that are extreme in intensity, finish does not match like finish on other parts—requires refinishing.

Class C—Moderately Serious, Demerit Value 10

1. May possibly cause an operating failure of the unit in service, e.g., clearance between terminals insufficient.

2. Likely to cause trouble of a nature less serious than an operating failure, such as substandard performance, e.g., cable not adequately supported.

3. Likely to involve increased maintenance or decreased life, e.g., insulation frayed (major).

4. Major defects of finish or workmanship, e.g., finish conspicuously scratched, designation omitted or illegible.

Class D—Not Serious, Demerit Value 1

1. Will not affect operation, maintenance, or life of the unit in service (including minor deviations from engineering requirements), e.g., sleeving too short (minor), or wiring reverse between two terminals of a simple make or break contact combination.

2. Minor defects of appearance, finish, or workmanship.

Standard Quality

Obviously the mere finding of defects and assessing demerits does not provide any measure as to whether the sample inspected is satisfactory, or as to whether it is better or worse than what could be considered adequate for the needs of the system. To provide some yardstick against which the quality of the sample, as measured in demerits, can be compared, we have developed the "standard-quality level." This may be defined as the expected average value of demerits per unit. The actual value will of course vary for the varieties of product according to the complexity and the engineering requirement structure.

This average demerit value is determined by a check of the records of quality-assurance inspection during a base period in which the product in question has delivered normally satisfactory service. Normally, such a base period would be 3 years in length and would be characterized by absence of adverse reports from the field, either from a WE installation or the operating telephone companies.

Of course, no product repeatedly has demerits per month exactly equal to the standard quality level. It is therefore necessary to determine whether the deviation above (fewer than expected) or below (more demerits than expected) is significant. For this purpose, two sigma limits are used rather than the three normally employed in statistical quality control. However, the first time a product is between two and three sigmas below standard, it is called border line and no action is taken on the basis that such a condition could occur solely due to chance. If this occurs a second time within 6 months, or if quality of a product is below standard by more than three sigmas, it is considered significant, and corrective action must be taken to bring the quality of that product within the normal band range of two standard deviations from the expected quality level.

One other point might be raised here, namely, the determination of sigma values in terms of demerits as applied

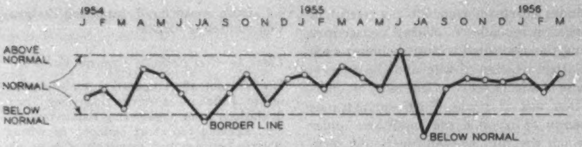


Fig. 1. Method of plotting border line and below normal points of a product

to the monthly rating. Sigma is determined by the formula $\sqrt{nC_s}$ when n is the monthly sample and C_s is standard-quality variance, which is further defined as a measure in the demerits per unit values to be expected of samples from product having quality at the standard-quality level. As in the case in the standard-quality level itself, the standard-quality variance will be different for each individual product to be rated.

Fig. 1 is a typical chart showing a product which is, in general, satisfactory and which indicates the method of plotting border-line and below normal points. Points above the upper two sigma limits obviously do not cause any concern for quality, but if long continued, might lead to an investigation to determine whether too much money is being spent to maintain quality that is better than what is expected or needed.

Publication of Reports on Quality

Completion of inspection and calculation of a rate would be of little aid in assuring the quality of manufactured product unless results are published and distributed so that persons in the proper positions of authority can take any requires action. For this purpose, two reports are issued. One, quarterly, shows a chart for every product which is checked and rated by the quality-assurance organization at all manufacturing locations, regardless of whether the quality may be normal or otherwise. This publication provides a historical record of the status of each product for a 3-year period and is particularly helpful in calling to the attention of supervision any quality trends which may manifest themselves over a period of time.

The second report is issued monthly and contains charts only on those products which quality-assurance inspection has shown to be below normal for the month. It also contains general indexes for the over-all performance of each plant and of the record of loose soldered connections found at each plant. This report is circulated at high management levels, and the appearance of a chart on a product in this publication is generally the signal for drastic corrective action on the part of the people responsible for its manufacture.

In addition to this function of maintaining a constant surveillance of quality, the quality-assurance organization has the right and the responsibility to halt immediate delivery of completed products in any case when sampling or other sources of information indicate that product quality is such that it should not be furnished to customers in the condition it was found by quality assurance.

Administration of the Quality Assurance

With the basic responsibility assigned to the quality-assurance organization, it is apparent that it must be free from any pressures for production. To accomplish this purpose, the organization has been set up under engineering rather than production supervision. At each manufacturing plant the quality-assurance resident head is not in any way responsible to the production supervision at that location. The entire organization is headed by a superintendent of quality assurance located at the company headquarters in New York, who reports through appropriate management to the vice-president of engineering for the company. At the manufacturing locations the amount of supervision will depend, of course, on the size of the plant and the kinds and the amount of the products manufactured there. Personnel include engineers, quality checkers, statistical clerks, secretaries, and possibly other clerical help.

Engineers have a wide range of responsibility. They design and sometimes build test equipment to help determine conformance of the product to basic engineering requirements. They investigate and prepare answers to complaints which may originate from installers or from customers. They analyze historical data and develop standard quality levels for new products or for those products that need a revision of their standard level.

Quality checkers perform the actual inspections using handbook instructions, drawings, and specifications applicable to the product and test instructions prepared by the quality-assurance engineer. They prepare the preliminary reports on the basis of results of their own inspection using the printed forms developed for the purpose. They do not, however, prepare any written instructions as to what requirements are to be checked or how any inspection is to be made. Statistical clerks each month calculate the actual quality rates using standard formulas and the preliminary reports prepared by quality checkers.

Summary

1. WE is both the principal manufacturer and the principal installer for the Bell System, turning a substantial portion of its product over to the customer in the form of working telephone installations.

2. Quality assurance at the completion of manufacture thus becomes a part of the larger over-all process; a transition between manufacturing and installation.

3. Inspection on a sampling basis is made to basic engineering requirements, and defects are assessed on a demerit basis according to seriousness.

4. Monthly rating compares a product with a standard-quality level, and deviations in excess of two sigmas are considered not normal.

5. Two reports are issued: one quarterly for all products, and one monthly for below normal products which require immediate corrective action.

6. Quality assurance is administered through an independent organization reporting to the vice-president and chief engineer. Personnel, in addition to supervisors, include engineers, quality checkers, statistical clerks, and other clerical help.

7. The discussion of quality assurance contained in this paper has been designed to give a general perspective of the over-all program employed by WE. Because of space and time limitations, it has not been possible to go into detail. Further information on the subject may be found in other periodicals.[1,2]

References

1. An Over-All Quality-Assurance Plan, G. D. Paterson. *Industrial Quality Control Magazine*, Rochester, N. Y., May 1956.

2. A Check Inspection and Demerit Plan, H. F. Dodge, M. N. Torrey. *Ibid.*, June 1956.

The Work of the International Telecommunication Union in the Field of Telephony

W. E. BLOECKER
MEMBER AIEE

G. A. CALDWELL
MEMBER AIEE

FEW ENGINEERS in North America are familiar with the work of the International Telecommunication Union or are even aware of its existence. This paper discusses the Union; its organization, how it functions, and what it is doing in the field of telephony. It also gives an assessment of the value of this work and an appraisal of its significance to the communication developments in the United States and Canada.

It is trite to say these days that the world is becoming smaller. Certainly telecommunications are helping make it so in the ease by which the expression of thought may be transmitted without distortion between nations and continents, and by the medium this provides for the achievement of ready understanding. As it affects the engineer, when such a medium becomes available to all men, he has to be aware of the influence of his thinking and his progress in achieving his objectives on his own continent, on his co-ordinates in other areas, and of their influence on him. The International Consultative Committee activities described here provide an instrument essential to this realization on a universal basis.

It is especially opportune to review this field of activity now, since, on January 1, 1957, the Consultative Committees on Telegraphy and Telephony were reorganized into a new single committee covering both fields.

Historical Background

The International Telecommunication Union is the existing symbol of international co-operation to provide a favorable climate for the development of the world-wide communication network as it is known today. It is a comparatively young body, having been established at Madrid, Spain, in 1932. Its ancestors go back much further, the International Telegraph Union dating from 1865 and the International Radiotelegraph Union from 1903.

The first seeds, however, were planted even earlier in central Europe; the first agreement on record, drawn by Prussia and Austria in the latter part of 1849, tied together the telegraph networks of those two countries. Within a few months similar bilateral agreements were made between Prussia and Saxony, and between Austria and Bavaria. All of these agreements classified and regulated telegraph traffic between the contracting governments. It is of interest that originally they were established to meet military and other governmental needs, and provided for public messages only to the extent and during the periods that facilities might be available.

The next logical step came within a year when these four states entered into the Austro-German Telegraph Union covering the mutual interchange of traffic. By the end of 1854 this group grew to include the Netherlands and four additional German states.

Meanwhile in western Europe, France concluded similar bilateral agreements with its neighbors Belgium, Switzerland, Sardinia, and Spain. These five states soon saw the advantages of a joint agreement and in 1855 created the Western European Telegraph Union.

It was probably inevitable that relations would be established between members of these two groups. Thus it was in 1852 that a third grouping developed when Prussia, Belgium, and France signed a Convention* regulating their mutual correspondence. By 1861 this group attracted as adherents 11 additional states.

The next big move was initiated by France in 1864 when all the major countries of Europe were invited to Paris. Twenty states responded. Out of this came the International Telegraph Union whose Convention was signed in 1865.

This union enjoyed an ever-widening sphere of influence. Asia Minor was included in 1863, Persia and India in 1868, Great Britain in 1872. At a meeting in 1875 Egypt was present under its own identity and Japan and the United States sent observers. By 1882 all areas of the world were represented, with the United States still in the role of observer.

Meanwhile the telephone had been invented in 1876, and by 1885 several international lines were in existence in Europe. At the Berlin (Germany) Telegraph Conference of 1885, the first steps were taken in the regulation of international telephone traffic.

By 1924, and especially in the period following the close of World War I, telephony had made great progress and was expanding at an increasing rate. In that year, in recognition of the need for an organization which would afford means for having meetings at which experts of various countries could consider the technical and operating problems of the European telephone system there was created the International Consultative Committee for Long Distance Telephony. In the 1925 Paris Conference of the International Telegraph Union, the Telephone Consultative Committee was brought into the Union as a semi-independent body. At the same meeting a similar Consultative committee for telegraph communications was established.

With the development of radio as a practical means of communication around the end of the 19th century, international co-operative action was initiated in the Berlin Conference of 1903 to work out problems to assure unhindered international radio communication. At that time safety of life at sea was a prime consideration. Nine countries, including the United States, attended. In a second Berlin Conference in 1906, an International Radiotelegraph Convention and Associated Radio Regulations were adopted by 29 countries.

The International Radiotelegraph Conference at Washington, D. C., in 1927, established the International Consultative Committee for Radio Communications, with similar aims as those already established for telegraphy and telephony.

As radio grew its political and economic effects became so great that no nation could afford to ignore them; by 1932 the International Radiotelegraph Union

---

* "Convention" as used here, and in international affairs generally, means a multiparty treaty to which governments are signatory.

Paper 57-785, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 27, 1957; made available for printing April 30, 1957.

W. E. Bloecker is with the American Telephone and Telegraph Company, New York, N. Y., and G. A. Caldwell is with the Bell Telephone Company of Canada, Ltd., Montreal, Que., Canada.

membership included almost every civilized nation on the globe.

The stage was now set for the final welding of world-wide communications under a single body. At Madrid in 1932, the 13th International Telegraph Conference and the fourth International Radiotelegraph Conference met simultaneously. A liaison was achieved for consideration of common questions; by joint meetings of the two Plenary Assemblies and by Joint Conference Committees. The most important achievement of the Madrid conference was the creation of a single Convention, containing the general principles common to the telegraph, telephone, and radio services. Thus was created the International Telecommunication Union.

The International Telecommunication Union

The International Telecommunication Union is recognized by the United Nations as the specialized agency responsible for taking such action as may be appropriate in the fulfillment of its purposes. Its fundamental basis of operation is the International Telecommunication Convention, the latest revision of which is that adopted at the Plenipotentiary Conference at Buenos Aires, Argentina, in 1952. This Convention was signed by delegates of 78 countries and territories of the world and subsequently ratified by practically all of the countries of the world.

To round out the provisions of the Convention, there are four sets of Administrative Regulations—Telegraph, Telephone, Radio, and Additional Radio—which are concerned with detailed relationships between correspondent countries in international communications. Many of the countries signatory to the Convention have officially indicated reservations regarding observance of one or more of these sets of regulations. These positions arise out of situations peculiar in those countries on international services. The United States and Canada are among those countries which have not adhered to the Telephone and Additional Radio Regulations because they include operating and rate matters based principally on European practices, and because these matters are the concern primarily of the operating agencies handling the communication services, rather than the governments themselves. This subject will be considered further in a later part of the paper.

MEMBERSHIP

Membership is open to any sovereign country and group of territories in the world, which meet certain general qualifications set down in the Convention.

Members are entitled to participate in conferences of the Union and are eligible to any permanent organ of the Union. Each member, large or small, has one vote in conferences and meetings.

PURPOSES OF THE UNION

The purposes of the Union are spelled out in the Convention and are quoted here in full:

1. The purposes of the Union are:
   (*a*) to maintain and extend international co-operation for the improvement and rational use of telecommunication of all kinds;
   (*b*) to promote the development of technical facilities and their most efficient operation with a view to improving the efficiency of telecommunication services, increasing their usefulness and making them, so far as possible, generally available to the public;
   (*c*) to harmonize the actions of nations in the attainment of those common ends.

2. To this end, the Union shall in particular:
   (*a*) effect allocation of the radio-frequency spectrum and registration of radio-frequency assignments in order to avoid harmful interference between radio stations of different countries;
   (*b*) foster collaboration among its Members and Associate Members with a view to the establishment of rates at levels as low as possible consistent with an efficient service and taking into account the necessity for maintaining independent financial administration of telecommunication on a sound basis;
   (*c*) promote the adoption of measures for ensuring the safety of life through the co-operation of telecommunication service;
   (*d*) undertake studies, formulate recommendations and collect and publish information on telecommunication matters for the benefit of all Members and Associate Members.

STRUCTURE OF THE UNION

The purposes of the Union form a fairly comprehensive directive. The magnitude of the task is pointed up by the character of the membership-peoples of the five continents of the world with their widely varying cultural, social, and economic development. To carry out these objectives calls for a well-considered organization which will provide for an amicable exchange of views, the resolution of differences, and the development of recommendations applicable and acceptable to the wide range of telecommunication needs of the members.

The organization must also be flexible enough to adapt itself to the changes in this rapidly changing world in which communications is such a vital part. This must be accomplished by a machinery with broad powers to act in the intervals between Plenipotentiary Conferences.

The present organization of the Union may be thought of as comprising two broad fields of endeavor. One of these is administrative, to assure that the purposes for which the Union was established are fruitfully being carried out; the other is technical, to provide for studies and the interchange of ideas relating to the engineering and operation of communication systems.

The organization of the Union is shown in Fig. 1. It consists of the following.

Plenipotentiary Conference

This is an international meeting of governments at the diplomatic level. It is the supreme organ of the Union. It provides general direction, establishes the budget and salary scales, elects members of the Administrative Council, revises the Convention as necessary, and makes agreements with other international organizations. It normally meets at 5-year intervals but the date can be changed with the consent of at least 20 members. (Its next meeting is now tentatively scheduled for 1959, after an interval of 7 years.)

Administrative Conference

This, too, is an international meeting at government level, normally convened at the same time and place as the Plenipotentiary Conference. Its primary function is to revise the Regulations annexed to the Convention. Provision is made for calling extraordinary and special administrative conferences with particular agenda.

Administrative Council

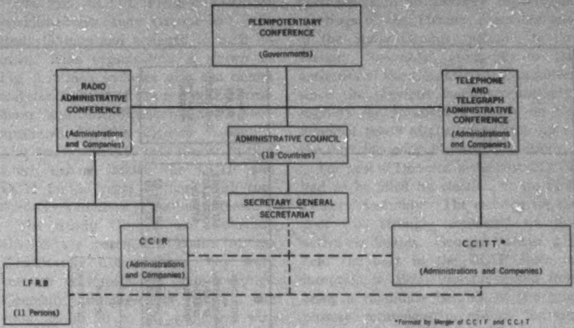
This body is composed of 18 members elected by the Plenipotentiary Conference. It may be considered as somewhat analogous to a board of directors. Meeting annually, its job is to see that the work of the Union proceeds smoothly within the boundaries of the Convention and the Regulations. It also has the power to conclude provisional agreements with other international bodies, subject to ratification at the following Plenipotentiary Conference.

General Secretariat

The duties of this organ are implicit in its title. It may be likened to the business office and general service departments of large companies.

International Frequency Registration Board

This provides an internationally recognized record of radio-frequency assignments made by the different countries, together with the date, purpose, and technical character of each assignment. It furnishes member countries advice with a view to the operation of the maximum practicable number of radio channels in those portions of the spectrum where harmful interference may occur.


PLENIPOTENTIARY CONFERENCE
(Governments)
RADIO ADMINISTRATIVE CONFERENCE
(Administrations and Companies)
TELEPHONE AND TELEGRAPH ADMINISTRATIVE CONFERENCE
(Administrations and Companies)
ADMINISTRATIVE COUNCIL
(18 Countries)
SECRETARY GENERAL
SECRETARIAT
CCIR
(Administrations and Companies)
CCITT*
(Administrations and Companies)
I.F.R.B
(11 Persons)

*Formed by Merger of C.C.I.F. and C.C.I.T. Effective January 1, 1957

Fig. 1. Organization of International Telecommunication Union

International Consultative Committee

The technical and operational features of communication services are the fields of study and recommendations of the International Consultative Committees. These bodies form the focal point of interest in this paper since their work is almost exclusively technical and professional.

The Buenos Aires Convention of 1952 continued the three Consultative Committees: Telegraph, CCIT (Comité Consultatif International Télégraphique); Telephone, CCIF (Comité Consultatif International Téléphonique); Radio, CCIR (Comité Consultatif International des Radiocommunications).

The question of combining the telegraph and telephone activities under a single consultative committee had long been debated. Following World War II, agitation for this move increased, and the 1952 Convention provided for the eventual merger of CCIF and CCIT.

Finally, in 1955, the members of the Union voted by telegraph ballot to merge the activities of CCIF and CCIT into a new International Telegraph and Telephone Consultative Committee to be known as CCITT (from the French equivalent of its name). This merger was effective as of January 1, 1957.

Membership in the Consultative Committees is available as a matter of right to members of the Union as prescribed in the convention.

From the discussion so far, it is evident that membership in the Union is reserved to governments of sovereign states and their territories. There are, however, large areas of the world, with well over 50% of the world's telephones, in which the operation of telephone services is conducted wholly or in part by private enterprise or a public corporation rather than by a department of the government. Included in this list are the United States, Canada, Mexico, Cuba, Chile, Spain, and Japan. In order to make it possible for such organizations to participate in the work of the Union for the good of all, provision is made for their respective governments to certify these private communication companies to membership in the Consultative Committees. These are known as Recognized Private Operating Agencies, and enjoy the full rights of membership, including, in the absence of any representative of their government, the right to vote. By invitation of the International Telephone Consultative Committee and with the concurrence of the State Department, the American Telephone and Telegraph Company has been a member of CCIF since 1929. In 1954 the Government of Canada accredited the Telephone Association of Canada (representing over 90% of the telephone development in the Dominion) as a Recognized Private Operating Agency to be a member of this Consultative Committee.

In many areas of the world the major portion of research and development of new systems and apparatus is carried out in the laboratories associated with equipment manufacturers. They obviously have much to contribute to the work of the Consultative Committees. In order to take advantage of this knowledge, representatives of manufacturers accredited by their governments, are granted membership as experts in the study groups charged with a wide variety of technical problems. They may participate under their own company's identity or as experts attached to the delegations of member administrations. In either capacity they make valuable contributions to the solution of problems which lead to the ultimate recommendations of the Study Group. However, since the Consultative Committees function in the interest of the operating communication services, both government and private, the manufacturers do not attend Plenary Meetings nor do they have the right to vote.

FINANCES OF THE UNION

Expenses of the Union which currently run in the order of about $\$1^1/_2$ million annually, are classified as ordinary and extraordinary. The former covers expenses of the Administrative Council, of the General Secretariat, of the International Frequency Registration Board, and of the Consultative Committees. They are borne by Members and Associate Members.

The extraordinary expenses are defined as all expenses pertaining to Plenipotentiary Conferences, Administrative Conferences, and meetings of the International Consultative Committees. These expenses are borne by participating governments, Recognized Private Operating Agencies, and scientific and industrial organizations which have agreed to participate in these conferences and meetings or which have actually participated.

Apportionment of the expenses of participants in the work of the Consultative Committee is decided on the basis of the classification chosen by the participant. These range from a 30-unit classification to one of 1/2 unit. The expense allocated to a particular member is determined by the ratio of the number of units in the class the participant has chosen to the number of units represented by the total participants.

The estimated total extraordinary expenditure of the CCITT based on Study Group activities for each of 1957

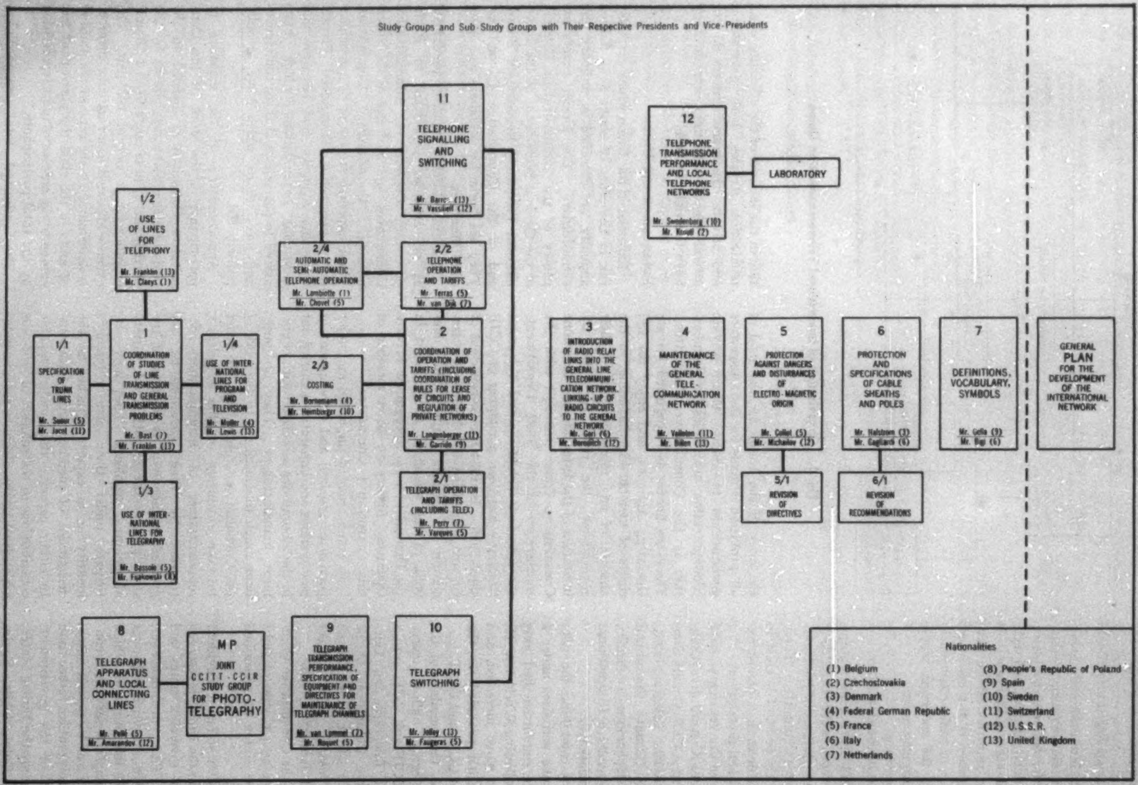

Study Groups and Sub-Study Groups with Their Respective Presidents and Vice-Presidents
1/1
SPECIFICATION OF TRUNK LINES
Mr. Sueur (5)
Mr. Jacot (11)
1/2
USE OF LINES FOR TELEPHONY
Mr. Franklin (13)
Mr. Claeys (1)
1
COORDINATION OF STUDIES OF LINE TRANSMISSION AND GENERAL TRANSMISSION PROBLEMS
Mr. Bast (7)
Mr. Franklin (13)
1/3
USE OF INTER-NATIONAL LINES FOR TELEGRAPHY
Mr. Bassole (5)
Mr. Fijalkowski (8)
1/4
USE OF INTER-NATIONAL LINES FOR PROGRAM AND TELEVISION
Mr. Muller (4)
Mr. Lewis (13)
11
TELEPHONE SIGNALLING AND SWITCHING
Mr. Barr (13)
Mr. Vassilieff (12)
2/4
AUTOMATIC AND SEMI-AUTOMATIC TELEPHONE OPERATION
Mr. Lambiotte (1)
Mr. Chovet (5)
2/2
TELEPHONE OPERATION AND TARIFFS
Mr. Terras (5)
Mr. van Dijk (7)
2/3
COSTING
Mr. Bornemann (4)
Mr. Heimburger (10)
2
COORDINATION OF OPERATION AND TARIFFS (INCLUDING COORDINATION OF RULES FOR LEASE OF CIRCUITS AND REGULATION OF PRIVATE NETWORKS)
Mr. Langenberger (11)
Mr. Garrido (9)
2/1
TELEGRAPH OPERATION AND TARIFFS (INCLUDING TELEX)
Mr. Perry (7)
Mr. Varques (5)
12
TELEPHONE TRANSMISSION PERFORMANCE AND LOCAL TELEPHONE NETWORKS
Mr. Swedenborg (10)
Mr. Kroutl (2)
LABORATORY
3
INTRODUCTION OF RADIO RELAY LINKS INTO THE GENERAL LINE TELECOMMUNICATION NETWORK. LINKING-UP OF RADIO CIRCUITS TO THE GENERAL NETWORK
Mr. Gori (6)
Mr. Borodich (12)
4
MAINTENANCE OF THE GENERAL TELE-COMMUNICATION NETWORK
Mr. Valloton (11)
Mr. Billen (13)
5
PROTECTION AGAINST DANGERS AND DISTURBANCES OF ELECTRO-MAGNETIC ORIGIN
Mr. Collet (5)
Mr. Michailov (12)
5/1
REVISION OF DIRECTIVES
6
PROTECTION AND SPECIFICATIONS OF CABLE SHEATHS AND POLES
Mr. Halstrom (3)
Mr. Gagliardi (6)
6/1
REVISION OF RECOMMENDATIONS
7
DEFINITIONS, VOCABULARY, SYMBOLS
Mr. Gella (9)
Mr. Bigi (6)
GENERAL PLAN FOR THE DEVELOPMENT OF THE INTERNATIONAL NETWORK
8
TELEGRAPH APPARATUS AND LOCAL CONNECTING LINES
Mr. Pellé (5)
Mr. Amarandov (12)
M P
JOINT CCITT-CCIR STUDY GROUP FOR PHOTO-TELEGRAPHY
9
TELEGRAPH TRANSMISSION PERFORMANCE, SPECIFICATION OF EQUIPMENT AND DIRECTIVES FOR MAINTENANCE OF TELEGRAPH CHANNELS
Mr. van Lommel (7)
Mr. Roquet (5)
10
TELEGRAPH SWITCHING
Mr. Jolley (13)
Mr. Faugeras (5)
Nationalities
(1) Belgium
(2) Czechoslovakia
(3) Denmark
(4) Federal German Republic
(5) France
(6) Italy
(7) Netherlands
(8) People's Republic of Poland
(9) Spain
(10) Sweden
(11) Switzerland
(12) U.S.S.R.
(13) United Kingdom

Fig. 2. Organization of the International Telegraph and Telephone Consultative Committee

and 1958 is about $50,000 to $60,000 and for 1959, the year of the next Plenary Assembly, about $100,000. The total number of units represented at the 1956 Plenary Assembly was 501¹/₂. Assuming approximately the same level in 1959, the expenses of a participant in the 30-unit class would be about $6,000 and in the 1/2-unit class about $100.

International Telegraph and Telephone Consultative Committee

With this background on the Union and the Technical Consultative Committees, we can now turn to the central theme of the paper, the field of telephony.

For over 30 years studies of technical and operating questions relating to international telephone service have been ably co-ordinated by the International Consultative Committee for Telephony, the CCIF. In the future its destiny will rest with CCITT, the newly created International Consultative Committee for Telegraphy and Telephony.

In December 1956 there were held at Geneva, Switzerland, the VIIIth and final Plenary Assembly of the CCIT, the XVIIIth and final Plenary Assembly of the CCIF, and the first Plenary Assembly of the CCITT. For the CCIF and CCIT these Plenary Assemblies concluded their respective affairs, and proposed a program for the future work of the new CCITT. For this new body, there were the tasks of organization, the election of a Director, and the establishment of a program of work and future meetings.

ORGANIZATION

Plenary Assembly

The Plenary Assembly is the supreme authority of the Consultative Committees. Among its duties are the election of a Director, the establishment of Study Groups and Substudy Groups to carry on its work, the approval of a work program, the approval of recommendations of the technical Study Groups, and the financial requirements of the Consultative Committee.

By provision of the Convention, the Plenary Assembly is normally held at 3-year intervals.

Study Groups

The real work of the Consultative Committee is accomplished in Study Groups, of which there is one for each of the broad areas of telephone and telegraph engineering and operations. In the larger and more complex areas, the work is divided among Substudy Groups.

The first CCITT Plenary Assembly established its Study Groups and Substudy Groups and assigned to each its tasks. This organization is shown in Fig. 2, which includes also the names and nationalities of the respective Presidents and Vice-Presidents who were appointed by the Plenary Assembly. This arrangement is very similar to that which existed under the CCIF and CCIT for 25 years or more, with consolidations in areas of mutual interest.

The Study Groups and Substudy Groups are permanent bodies of the Consultative Committee. They work through correspondence and meetings. Generally speaking, they range in size from about 20 to 30 members, with attendance at meetings of delegates and experts ranging from 30 or so to well over 100. The larger groups would obviously labor under extreme difficulties in trying to resolve complex questions, such as the design parameters of a coaxial cable system. To handle such problems, Working Groups composed of a small number of well-qualified experts are created. These, set up for specific limited tasks, are temporary, and are disbanded when they have reported their findings back to the Study Group or Substudy Group which had created them.

Language

By the provisions of the Convention the official languages of the Union are English, French, Spanish, Russian, and Chinese. In the smaller Study Groups or Working Party meetings, agreement is usually obtained to conduct the work in French and English with consecutive translation. While this might appear cumbersome, it actually works quite well due in larger measure to the competence of the interpreters and their knowledge of the technical terminology of telephony.

For the 1956 Plenary Assembly at Geneva and the larger study groups meeting immediately preceding it, simultaneous translation facilities were made available. In these meetings the proceedings were conducted by agreement in English, French, Spanish, and Russian. This did not seem to cause any particular difficulty for the Chinese delegates who were well versed in English.

The Director

The Director elected by the Plenary Assembly co-ordinates the work of the Consultative Committee, including its Plenary Assembly and Study Groups, and is responsible for the documents of the committee.

He is assisted by the Secretariat. He participates in an advisory capacity in meetings of the Plenary Assembly and of the Study Groups; submits to the Plenary Assembly a report on the activities of the Consultative Committee since the previous meeting; and also submits for approval a report on the financial needs of the committee up to the next following Plenary Assembly.

The post of Director of the new CCITT had to be filled by election at its First Plenary Assembly. The election for the post was strongly contested by five active candidates. Georges Valensi, who held the post in the CCIF since its inception in 1924, was prevented from being a candidate because of the compulsory retirement provisions for employees of the Union.

The light in which he is regarded was expressed at the final Plenary Assembly of the CCIF by the chief delegate of Italy, Mr. Gneme, and by the delegate of the United Kingdom, Mr. Barron. Extracts from their remarks are cited here because they serve to evaluate the importance of the Director's role as well as the highly deserved tribute they pay to Mr. Valensi. Mr. Gneme said in part:

> As regards Mr. Valensi's decisive contribution to the development of the international telephone network, I would recall that the Committee and its Laboratory, once set up, speedily produced fruitful results. In 1924, it was hardly possible to put a call through from one country to the next, . . . by 1938, all the European countries were linked by high-quality lines, ...
>
> After this initial period of creative activity, the problem arose of setting up new international lines with a higher transmission speed than the old coil-loaded cables. Thanks to the use of carrier systems, a great saving was effected in raw materials at the same time as quality was improved. The second modern long-distance line network has been set up in the light of the investigations undertaken by the CCIF, and the recommendations it has issued, during the last few years. The public has been made to benefit by technical inventions as soon as they got past the laboratory stage.

Mr. Barron said in part:

> Mr. Valensi has the exceptional ability of being able to carry out all his duties, both great and small, with the same abounding energy and impressive efficiency. The CCIF owes much of its success to this fact, because clearly the successful direction of such a body with its many complications and its wide technical span, must make heavy demands on its director, and the CCIF can count itself outstandingly fortunate in having had the services, over such a long and vital period, of a man who is at once an engineer, and administrator, and a politician (in the best sense). A man who can converse freely and fluently in a

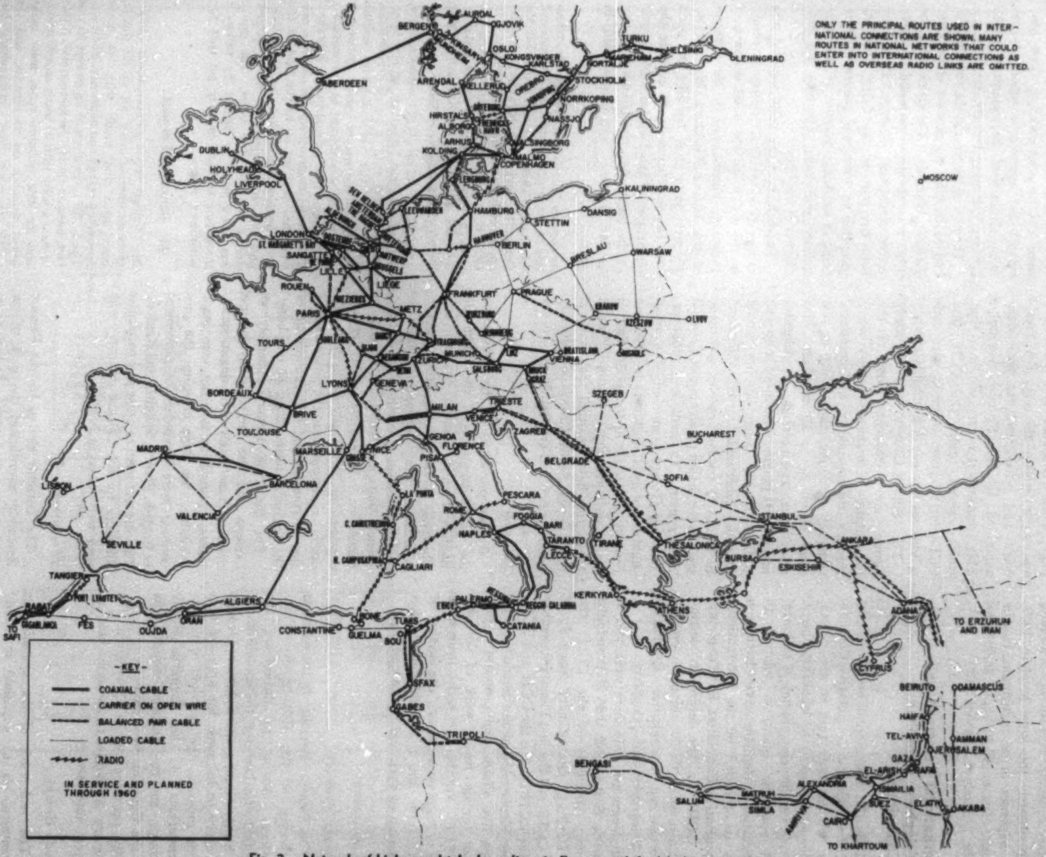

ONLY THE PRINCIPAL ROUTES USED IN INTERNATIONAL CONNECTIONS ARE SHOWN. MANY ROUTES IN NATIONAL NETWORKS THAT COULD ENTER INTO INTERNATIONAL CONNECTIONS AS WELL AS OVERSEAS RADIO LINKS ARE OMITTED.
-KEY-
COAXIAL CABLE
CARRIER ON OPEN WIRE
BALANCED PAIR CABLE
LOADED CABLE
RADIO
IN SERVICE AND PLANNED THROUGH 1960
BERGEN
GJOVIK
OSLO
KONGSVINGER
KARLSTAD
TURKU
HELSINKI
LENINGRAD
NORTALJE
STOCKHOLM
NORRKOPING
NASSJO
HELSINGBORG
ABERDEEN
ARENDAL
HIRSTALS
ALBORG
ARHUS
KOLDING
MALMO
COPENHAGEN
KALININGRAD
DANSIG
MOSCOW
DUBLIN
HOLYHEAD
LIVERPOOL
HAMBURG
STETTIN
BERLIN
WARSAW
BRESLAU
LONDON
ST. MARGARET'S BAY
SANGATTE
LILLE
LIEGE
ROUEN
PARIS
METZ
FRANKFURT
PRAGUE
LVOV
TOURS
STRASBOURG
MUNICH
LINZ
VIENNA
BRATISLAVA
ZURICH
SALZBURG
GRAZ
LYONS
GENEVA
BORDEAUX
BRIVE
MILAN
TRIESTE
VENICE
SZEGED
TOULOUSE
MARSEILLE
NICE
GENOA
FLORENCE
PISA
ZAGREB
BUCHAREST
BELGRADE
SOFIA
MADRID
BARCELONA
LISBON
VALENCIA
SEVILLE
PESCARA
ROME
NAPLES
FOGGIA
BARI
TARANTO
LECCE
TIRANE
ISTANBUL
THESALONICA
ANKARA
BURSA
ESKISEHIR
C. CARBONARA
CAGLIARI
TANGIER
PORT LYAUTEY
RABAT
CASABLANCA
TO SAFI
FES
OUJDA
ORAN
ALGIERS
CONSTANTINE
BONE
GUELMA
TUNIS
PALERMO
REGGIO CALABRIA
CATANIA
KERKYRA
ATHENS
ADANA
TO ERZURUM AND IRAN
CYPRUS
SFAX
GABES
TRIPOLI
BENGASI
BEIRUTO
DAMASCUS
HAIFA
TEL-AVIV
AMMAN
JERUSALEM
GAZA
EL-ARISH
ISMAILIA
ALEXANDRIA
SALUM
MATRUH
SIMLA
SUEZ
ELATH
AKABA
CAIRO
TO KHARTOUM

Fig. 3. Network of high-speed telephone lines in Europe and the Mediterranean Basin

number of languages, and who couples with his efficiency those vital qualities in human relationships of humor, understanding, mystery, and humanity.

The position of Director is thus regarded properly as a key position in the field of international communications. The election finally went on the fourth ballot to Jean R. Rouvière of France. He resigned as Director General of Telecommunications for France to assume his new duties.

Presidents and Vice-Presidents of Study Groups and Substudy Groups

These, too, are positions of considerable influence. The appointments in the new CCITT were made from European administrations and the United Kingdom with one exception: a delegate of Pakistan was appointed Vice-President of the Study Group for the general plan for the development of the international network of Europe, the Mediterranean Basin, and South Asia.

This situation reflects the long history of emphasis primarily on the problems of the European network. In recent years, however, the deliberations of the Study Groups are becoming increasingly of world-wide significance. In the future, therefore, it can be expected that outstanding engineers from other areas of the world will find their way into the posts of President and Vice-President of the Study Groups.

It is the job of the Presidents to direct the efforts of the Study Groups to the solution of the problems presented to them. Considerable flexibility is needed in reconciling conflicting points of views, influenced perhaps at times by existing equipment situations, or established operating procedures. A knowledge of such conditions is evidently of great value to the Presidents. Some facility of expression in both French and English is also helpful for these positions, but is not essential.

METHODS AND PROCEDURES

Any member of the Consultative Committee can elect to participate in the work of any of the Study Groups and should, of course, do so where it is felt the country's interests will be served by such participation, or where it can make significant contributions.

The work of the Study Groups consists of consideration of questions which are of active interest and significance to the current developments of the members themselves. Authority to consider these questions is given principally by the Plenary Assembly.

A typical program is that studied by CCIF in the period 1954–56. It consisted of the following:

Subject	Number of Questions
Transmission lines and systems	61
Transmission performance of telephone sets and local networks, etc.	18
Switching and signaling	18
Electrical protection and corrosion	43
Traffic operations and rates	25
Integration of wire and radio systems	19
Maintenance of lines and systems	16
Definitions, vocabulary, symbols	7

Obviously, complete agreement on a program as comprehensive as this is not possible. It frequently happens that questions are continued from one period to the next, with the same or new wording depending on current developments in the particular field.

Normally the questions for study are forwarded to the members of the Study Groups soon after the Plenary Assembly which authorized them. Many questions bracket the areas of interest of two or more Study Groups and are jointly studied. Any responses or contributions which the members can make are due approximately a year in advance of the next Plenary Assembly. There is thus a period of about 2 years in which to study the questions and propose solutions. It is to be expected that because of the variations in interest of the communication problems under study, and of limitations in manpower, all members will not respond to all questions; in fact, no one member ever has.

During this same period, a Working Party assigned specific tasks will have had exchanges of correspondence, and probably one or two meetings to prepare a report on its activities.

Some months, or even as much as a year, preceding the Plenary Assembly, the Study Groups may meet to consider all the responses received, together with the reports of any Working Parties they had created. Out of these meetings come draft recommendations, and decisions as to whether questions have been completed or should be continued. Normally as many as half the questions are of a fundamental nature or deal with a rapidly developing aspect of telephony, and may be continued in study over a long period of years. For such, interim recommendations, which have been based on the current state of the art, are drafted.

Additional meetings of Study Groups, primarily to review and edit the draft recommendations, are held immediately prior to the Plenary Assembly. The final stamp of approval of recommendations is made by the Plenary Assembly, which also makes provision for their publication.

The approved recommendations aspire to be a common acceptable basis for establishing international circuits or for engineering internal networks which are interconnected with such circuits. These recommendations are not mandatory and do not carry the force of regulations. They do stand as a guide and objective for all to follow as the state of the art and economics permit. In fact, the soundness of these recommendations and their adaptability to various situations tend to make it unnecessary to establish rigid regulations in these matters.

For example, the CCITT 1956 Plenary Assembly inherited from CCIF, and continued in study, a question relating to the design practices and objectives used by its members for their national trunk and exchange networks. The presence of this question on the agenda reflects recognition of the fact that as continuing improvements are made in international circuits, the limiting factor on the over-all grade of service becomes in an increasing degree the performance of the national facilities involved in completing the call. Generally, it might be expected that national networks providing a modern grade of service based on forward-looking engineering, are automatically adequate for completion of international calls. Unfortunately, for a variety of reasons, and some very good ones at that, this is not universally true. The responses to this question will be of great value in pointing up the need for common objectives on a world-wide basis.

ACHIEVEMENTS

It would be impractical to list completely the accomplishments of the International Telephone Consultative Committee during its lifetime.* As a matter of fact, the recommendations resulting from its work, useful in current operations, fill six volumes totalling some 1,500 pages. These are supplemented by voluminous annexes which contain documentation in support of the recommendations. It will be useful, however, to outline the broad areas of activity to indicate the wide scope of its activities.

Since the Committee's primary concern is good international telephone service (fast, easy to use, dependable, and at reasonable cost), one of its early accomplishments was the establishment of

---

* The first Plenary Assembly of the CCITT adopted a recommendation by the CCIF to organize a task force to prepare, in French and English, a complete history of the CCIF as a publication of the Union.

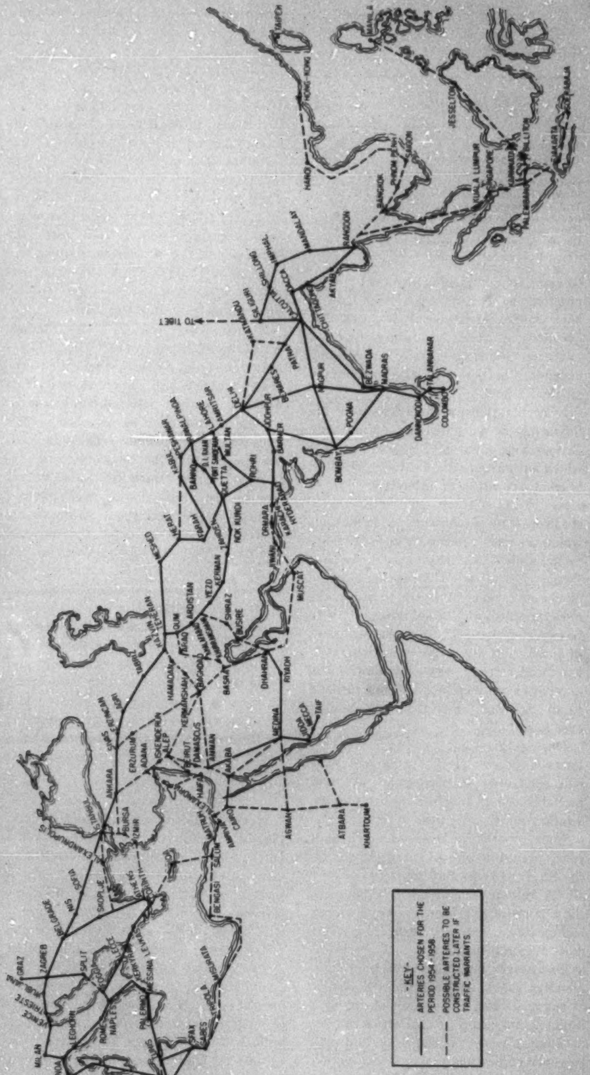

-KEY-
— ARTERIES CHOSEN FOR THE PERIOD 1954-1958
- - - POSSIBLE ARTERIES TO BE CONSTRUCTED LATER IF TRAFFIC WARRANTS
TAIPEH
MANILA
HONG-KONG
JESSELTON
HANOI
PHNOM PENH
SAIGON
BANGKOK
KUALA LUMPUR
SINGAPORE
PALEMBANG
DJAKARTA
MANDALAY
RANGOON
AKYAB
CHITTAGONG
DACCA
CALCUTTA
SILIGURI
KATMANDU
TO TIBET
PATNA
BENARES
NAGPUR
BEZWADA
MADRAS
COLOMBO
POONA
BOMBAY
DELHI
JODHPUR
BARMER
AMRITSAR
LAHORE
MULTAN
ROHRI
QUETTA
KARACHI
HYDERABAD
ORMARA
KABUL
PESHAWAR
HERAT
MESHED
KERMAN
YEZD
ARDISTAN
QUM
TEHERAN
KAZVIN
TABRIZ
HAMADAN
KERMANSHAH
BAGHDAD
BASRA
BUSHIRE
SHIRAZ
MUSCAT
JIWANI
DHAHRAN
RIYADH
MEDINA
JIDDA
MECCA
TAIF
AMMAN
DAMASCUS
BEIRUT
ALEP
ADANA
ERZURUM
ANKARA
ISTANBUL
BURSA
IZMIR
ATHENS
SOFIA
NIS
BELGRADE
SKOPLJE
ZAGREB
GRAZ
LJUBLJANA
TRIESTE
VENICE
MILAN
GENOA
MENTON
CANNES
LEGHORN
ROME
NAPLES
PALERMO
MESSINA
SPLIT
TUNIS
SFAX
GABES
TRIPOLI
MISURATA
BENGASI
SALUM
CAIRO
HAIFA
AGWAN
ATBARA
KHARTOUM
BONE
CONSTANTINE
GUELMA

Fig. 4. Network of high-speed telephone lines in South Asia and the Middle East and interconnection with the Mediterranean Basin

transmission design objectives for international circuits. This led logically to agreements on the design parameters for transmission systems of all types, carrier and radio. This is a matter of paramount importance, when it is realized that an international call in Europe, going no farther than one from Boston, Mass., to Washington, D. C., or Quebec to Ottawa, in Canada, may be handled by carrier systems where the ownership of the terminals will be divided between two nations, the terminals themselves may have been actually supplied by two different manufacturers, and the system may have been routed over the lines of still a third country.

Concurrently with the establishment of design objectives, agreement had to be reached on methods of measurement to assure that the objectives were being met. This, in itself, has covered virtually every aspect of electrical communication measurements.

Dependability of service rests not only on proper design, but is equally dependent on the skill and thoroughness with which the facilities are maintained and adjusted. Activity in this field has produced agreement on maintenance methods, routines for establishing, and periodic testing of circuits, etc.

As already discussed, no matter how well designed and maintained the international network may be, the grade of service will be determined to a very important degree by the characteristics of the telephone instruments at the subscriber's premises, and of the connecting facilities of the national networks. The telephone sets of the world, supplied by a great number of manufacturers, have probably the widest divergence in their essential transmission characteristics. A great deal of effort has been expended in trying to arrive at a common basis of comparison which would permit a generally accepted measure of the quality of performance of the individual sets. It cannot be said at this time that this matter has been completely resolved.

Many of the nations of Europe have made extensive progress in operator and direct distance dialing within their own borders. This inevitably has led to studies for the extension of dialing, both by operators and by the subscribers direct on international calls. This is another problem of great complexity because it means the integration of dialing and signaling systems developed for the national networks, into the international network. As would be expected of any progressive and free-thinking organization, many plans were offered. After some years of study two plans seemed to offer the best possibilities. Both arrange for supervision and dialing information to be transmitted in the voice band; one uses two frequencies to accomplish this, the other uses one. These were given extensive service trials in two areas of the European continent and some final recommendations regarding both have now been established.

In order to demonstrate what has been accomplished in international dialing, experimental arrangements were set up so that delegates at the meetings in Geneva in December 1956 could dial directly from telephones at the conference building into the networks of Brussels, Belgium; Germany; London, England; Milan, Italy; the Netherlands; Paris, France; and Stockholm, Sweden.

The accomplishments so far discussed are purely technical. Building on this work, the Committee has also produced a fundamental switching plan for international service. Initially this was confined to the continent of Europe and the United Kingdom. As the second step, the plan was extended to include the countries bordering on the Mediterranean Sea, and the third step was to make tentative plans for extensions to the Middle East and South Asia. Fig. 3 shows the network of high-velocity facilities (in service and proposed) in Europe and the Mediterranean Basin, and Fig. 4 the projected main arteries to the Middle East and South Asia.

The operation of such networks, of course, involves agreement on operating procedures, priority of calls, methods of billing, and apportionment of revenues, this last a difficult problem in connection with alternate routing through third countries.

The Consultative Committee has also concerned itself with requirements for program circuits used in connection with radiobroadcasting; the special considerations when voice-frequency telegraph circuits are operated on telephone-type channels; the interconnection of mobile radio systems with the wire network; and the very special types of problems associated with the long intercontinental circuits, which now span the seven seas.

The Committee has given a great deal of attention to problems of external interference and considerations of safety. This has resulted in setting specific objectives together with recommended practices in dealing with problems of interference and hazards to communication systems. The Committee's recommendations, based upon the pool of general experience, have been constructive.

Impressive as the accomplishments of this Consultative Committee are, it still has a formidable program ahead of it. To mention a few of its problems there are a rating system for the transmission performance of telephone sets; the transmission requirements for facilities used for television transmission; the contemplated extension of the standard European 6-megacycle coaxial system to 12 megacycles; the integration of radio systems, and particularly microwave, into the telephone network; and a reconsideration of the transmission design and maintenance of international circuits in the European network in view of the widespread availability of high-velocity facilities. (It has appeared to forward-looking European engineers a bit incongruous that a New York–Copenhagen trunk via the new transatlantic cable, about 5,000 statute miles, is operated at 6 decibels, whereas a 290-mile London–Paris call at present is operated at 7 decibels.)

Publications

It was mentioned earlier that the recommendations of the Telephone Consultative Committee have been published in a series of six books supplemented by other volumes containing related documentation furnished by the member organizations. In addition, there are special volumes covering the general switching plan, semi- and full automatic operation of the European network, directives on corrosion and protection, operators' instructions, maintenance phrases, definitions, and vocabulary. These volumes are more than bare recommendations. They include extensive background material, affording a better appreciation of the factors leading to the recommendations. This has proved to be of particular value to those countries whose communication systems are in the early stages of development.

Technical Assistance to Underdeveloped Countries

The Consultative Committees have made significant contributions to the United Nations expanded program for technical assistance to underdeveloped areas. The publications of the Telephone Consultative Committee have been an important part of this activity. Going beyond this, however, the Committee, through its acquaintanceship with engineers all over the world, and its knowledge of their particular qualifications, has been able to suggest candidates to the United Nations for specific missions to countries desiring aid. It has also been in a position, through its work, to make comments on the soundness of plans proposed by some of these countries, and has suggested modifications which might result in a more effective use of each country's resources. In another field, it has been able to give advice in connection with locating students receiving fellowship grants in specific areas. The Plenary Assembly in Geneva in 1956 reaffirmed this program of aid, and in addition concurred in the creation of a joint CCITT-CCIR (Radio) Working Party to examine further what additional contribution could be made by the Consultative Committees to the Technical Assistance Program.

In the discussions directed towards the Consultative Committees' efforts in the technical assistance program there seemed to be a strong feeling of concern on the part of the underdeveloped countries with programs for expanding their telecommunications, lest they be misguided into the purchase of systems which would not be compatible with other internal or external systems. The desire of such countries to obtain access to universal communications by facilities in keeping with the recommendations of the Consultative Committee was evident. The Plenary Assembly, by formal declaration, urged the underdeveloped countries to avail themselves of the full advantages of active membership of the Consultative Committee and its Study Groups.

Some Considerations of Significance

To the engineer it is evident that one of the principal achievements of the Consultative Committee lies in the awareness of a common objective which it has succeeded in creating against a background of diverse national, political, and traditional points of view.

Regarding communications in its worldwide aspect, it may naturally be asked why solutions which have been proved in application in North America do not always find ready and complete acceptance in Europe and vice versa. This, of course, points up the fact that sound engineering is not a technical matter alone, but must inevitably weigh also questions of investment, operating costs, climatic conditions, manufacturing, availability of manpower, money situations, political factors, and others. There is the basic need to make available universally the fundamental concepts and techniques of improved communications which may be applied by any nation in the light of its own national circumstances. The Consultative Committee provides the medium and the knowledge for such a constructive approach.

It is of interest, and probably of value, to remark that communication developments, such as the concept of carrier over both physical and radio channels, have inevitably resulted in a changing point of view, from one of restricted use of international media, both by regulation and by cost of service, to one of abundance producing readier access and lower costs. In this regard it may be observed that, from the authors' point of view, the Telephone Regulations for the European region (Paris revision 1949) annexed to the International Telecommunication Convention, are in many respects already outmoded in such features as classification of calls and priorities of acceptance of business offered to the international network. This is particularly true in view of present-day concepts of direct distance dialing.

This brings out an apparent difference in attitude on such matters as regulations which could affect the day-to-day operations of a communication network. On the North American continent such regulations, which can be changed only by action at the diplomatic level, are considered undesirable by the engineer because of the restrictions they may well place on taking prompt advantage of advances in the art resulting from improved understanding and research. On the European continent, however, it would appear that regulation between nations is accepted as a means to ensure adherence to standards in design or operation which have been debated to the point of common acceptance. In the operation of the communication networks of North America, such a position appears inflexible and forbidding. Further, on this continent, the operating agencies themselves have the responsibility for determining detailed practices which they will follow within the scope of broad governmental regulation.

These considerations, together with the fact that the Regulations are subscribed to by the countries concerned at the diplomatic level, and call for their imposition by their governments, makes adoption of them in North America all the more unacceptable. Consequently, many countries, including the United States and Canada, in the Final Protocol to the 1952 Convention, took official formal exception to the imposition of Telephone Regulations as an obligatory matter.

Significance of the Consultative Committee's Work to North America

The relations between the United States and Canada in the field of telephony are such that there has developed a mutual understanding, and a working together towards common objectives in telephone service quality. Major telephone operating organizations of Canada and the United States interchange information and experience freely, and improvements are used as soon as practicable after they appear. In light of this situation and of the relatively small part intercontinental traffic has so far played in the over-all operations, together with the highly special character of facilities to handle it, there has not been an urgency for establishing a close working liaison with the methods and designs of European systems. As a result, the Recognized Private Operating Agencies of the United States and Canada have not considered it appropriate to take a position on decisions relating strictly to the European region. They have, however, taken a very active part in the deliberations on truly worldwide questions.

This is not to imply a passive role. A member since 1929, the American Telephone and Telegraph Company, by virtue of its scope and diversity of operations, and its technical advances, has been in a position to contribute importantly to all phases of CCIF work. The Telephone Association of Canada, while a comparative newcomer, is already making its influence felt. On the other hand, both organizations are deriving benefit from their activity in the international body, not the least of which is the mutual respect of engineers on both sides of the Atlantic.

But new conditions are changing the picture. Although intercontinental radiotelephone service has been available for 30 years, the upsurge in traffic which followed the placing in service of the transatlantic cable, has made it evident that a new era in intercontinental telephony has been born. It is going to be necessary that the objectives in transmission design, and in operation of systems of telephony, be made compatible to the degree that the advantages offered by these intercontinental circuits are fully utilized in the interest of better communications. This can be done only by world-wide recognition of the importance of good objectives in the design of all national networks which it will be practicable to interconnect. The North American continent is thus inevitably being drawn into deeper participation in broader areas of the Consultative Committee's work to the end of gaining universal recognition of objectives, and of the necessity of constant planning to meet them both in Europe and on this continent.

In this paper, attention has been directed to the field of international telephony. There are equally stimulating and impressive stories to be written of the work of the International Consultative Committees for Telegraph and for Radio Communication. It may not be amiss to point out that all of these International Consultative Committees in making their outstanding contributions to human welfare, have been directed and influenced in all their decisions by the thinking of engineers. The instinctive desire of this profession to obtain the best practical solution in a sea of difficulties stands out here, independent of influences that may be prejudiced by nationally restricted points of view. In this there may indeed be a lesson for meditation.

References

1. INTERNATIONAL TELECOMMUNICATION CONVENTION, BUENOS AIRES, 1952. General Secretariat, International Telecommunication Union, Geneva, Switzerland.

2. THE INTERNATIONAL TELECOMMUNICATION UNION (book), G. A. Codding, Jr. E. J. Brill, Leiden, Netherlands, 1952.

Equations for the Inductances and Current Distribution of Multiple-Conductor Single-Phase and Polyphase Busses

T. H. CHIN
ASSOCIATE MEMBER AIEE

THOMAS J. HIGGINS
FELLOW AIEE

THE purpose of this paper is twofold: to derive general equations for calculating the currents and inductances associated with each subconductor of a multiple-conductor single-phase or polyphase bus; and to illustrate application of these equations by numerical analysis of several typical busses.

The essential values of the general equations may be epitomized as follows. General equations for calculation of the inductances and currents associated with the individual phases of multiple-conductor busses comprised of conductors of structural shape are available in the literature.[1] However, as exhaustive survey[2] reveals, general expressions for these quantities as associated with the individual conductors of the phases are not available, except for the special case of the bus comprised of coplanar phases, each comprised of two channels placed flange to flange, a geometry detailed by Siegel and Higgins.[3] But under the heavy loads typical of modern central station, electric furnace, electric welder, and other industrial plant-distribution systems, there has occurred an increasing use of busses comprised of multiple-conductor phases. Accordingly, as stressed by Schurig,[4–6] Jones (in a private letter to the second author), and other designers of such busses, there is a pressing need of equations enabling calculation of the aforementioned electrical quantities in the course of design. Precisely such equations are advanced in this paper.

The scope of this paper is to be summarized thus. By use of well-established basic theory, specific general equations for the inductance and current associated with the individual subconductors of a multiple-conductor bus are effected through use of geometric-mean-distance theory. With these available, application is illustrated by specific calculation of the inductances, currents, and voltage drops on the subconductors of three specifically dimensioned multiple-strap busses investigated experimentally by Schurig.[4–6] Finally, confirmative of the correctness of both the general equations and of the numerical calculation, and corresponding calculated and measured values of reactance are found to be in excellent agreement.

Derivation of Basic Equations for Individual Conductors

INDUCTANCE

From well-established theory[1] the inductance associated with the kth conductor of a bus comprised of N long parallel nonmagnetic conductors, each having a uniform but otherwise arbitrary cross section, is

$$L_k=-2\sum_{i=1}^{N}(I_i/I_k)\log D_{ik}=\sum_{i=1}^{n}L_{ik}$$

abhenrys per cm (centimeter) of conductor length (1)

where D_{ik} is the geometric mean distance between the ith and kth conductors, and I_i is the current (expressed as a complex number) associated with the ith conductor.

By Kirchhoff's current law, the sum of the currents is

$$I_a+I_b+I_c+\ldots+I_N=0 \tag{2}$$

If $S_1, S_2\ldots, S_m$ are the components of a total area S; if $D_1, D_2, \ldots, D_m$ are their respective geometric mean distances from a second area S'; and if D is the geometric mean distance between S and S'; then

$$SS'\log D=S_1S'\log D_1+S_2S'\log D_2+\ldots+S_mS'\log D_m \tag{3}$$

Let S' be comprised of component areas $S_1', S_2', \ldots, S_n'$. Then repeated use of equation 3 to obtain the geometric mean distances between the components S_i $(i=1,\ldots,m)$ and $S_j'(j=1,\ldots,n)$ yields

$$SS'\log D=\sum_{i=1}^{m}\sum_{j=1}^{n}S_iS_j\log D_{ij} \tag{4}$$

Alternatively, if the components of S' are numbered from $m+1$ to $m+n$

$$SS'\log D=\sum_{i=1}^{m}\sum_{j=m+1}^{m+n}S_iS_j\log D_{ij} \tag{5}$$

Equation 4 is preferable for determining the self geometric mean distance of an area to itself. For determining the geometric mean distance between two distinct areas equation 5 is preferable.

Next, from fundamental electromagnetic principles,[1] if S_a and S_b are the cross-sectional areas of a single-phase circuit carrying a current I, the vector potential $A(X,Y)$ at a point (X,Y) is

$$A(X,Y)=\text{constant}-2\iint_{S_a}w_a\log(|X-x|^2+|Y-y|^2)^{1/2}dxdy-2\iint_{S_b}w_b\log(|X-x|^2+|Y-y|^2)^{1/2}dxdy \tag{6}$$

where w_a and w_b are, respectively, the current densities in S_a and S_b, taken positive for current in one direction and negative for current in the other direction. The energy W in the magnetic field associated with the circuit is

$$\begin{aligned}W&=(1/2)\iint_{S_a}A(X,Y)w_a dXdY+(1/2)\iint_{S_b}A(X,Y)w_b dXdY\\&=-(2w_aw_bS_aS_b\log D_{ab}+w_a^2w_b^2\log D_{aa}+w_b^2S_b^2\log D_{bb})\\&=I^2\log(D_{ab}^2/D_{aa}D_{bb})\end{aligned} \tag{7}$$

Finally, as $W=LI^2/2$, it follows from equation 7 that

$$L=2\log D_{ab}^2/D_{aa}D_{bb} \tag{8}$$

If each of the two conductors is comprised of several subconductors connected in parallel, m comprising that of cross section S_a and n comprising that of cross section S_b, substitution from equation 4 and equation 5 in equation 8 yields the following

$$L=-(2/I^2)\sum_{i=1}^{m+n}\sum_{j=1}^{m+n}w_iw_jS_iS_j\log D_{ij} \tag{9}$$

SHORT-CIRCUIT FORCES

Equation 9 for the inductance enables easy derivation of an equation yielding the short-circuit force F_j on the jth conductor of a bus comprised of N conductors. From basic theory[1]

$$F_{jx}=\partial W/\partial x_j;\quad F_{jy}=\partial W/\partial y_j \tag{10}$$

wherein x_j and y_j define the position of

Paper 57-645, recommended by the AIEE Basic Sciences Committee and the Industrial Power Systems Committee of the Power Apparatus and Systems Division and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 27, 1957; made available for printing June 11, 1957.

T. H. CHIN is with Carnegie Institute of Technology, Pittsburgh, Pa., and THOMAS J. HIGGINS is with the University of Wisconsin, Madison, Wis.

This paper is based on a thesis supervised by Prof. Higgins and submitted by Prof. Chin to the faculty of the University of Wisconsin in partial fulfillment of the requirements for the degree of Master of Science in electrical engineering.

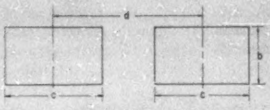


Fig. 1. Cross section of arbitrarily dimensioned rectangular-conductor bus: width of conductor greater than height

the jth conductor in a system of Cartesian co-ordinates, x and y, in a plane of the cross section of the bus; and F_{xj} and F_{jy} are the x and y components, respectively, of the total force F_j on the jth conductor. W, the energy in the magnetic field of the bus, is given by equation 11

$$W=\sum_{i=1}^{N} L_k I_k^2/2 \tag{11}$$

where L_k is the inductance associated with each conductor, as given by equation 1. If the conductors are subdivided and if the current density is uniform over each conductor cross section

$$L_k=-2\sum_{i=1}^{N}(I_i/I_k S_i S_k)\sum_{m=1}^{p_i}\sum_{n=1}^{q_k} S_m' S_n' \log D_{mn} \tag{12}$$

where $S'_1, S'_2, \ldots, S_{p_i}'$, are the p_i components of a total area S_i; $S_1', S_2', \ldots, S_{q_k}'$ are the q_k components of a total area S_k; and D_{mn} is the geometric mean distance between the mth component of S_i and the nth component of S_k. Then substituting equation 11 in equation 10 yields

$$F_{jx}=\sum_{k=1}^{N}(I_k^2/2)\partial L_k/\partial x_j \tag{13}$$

$$F_{jy}=\sum_{k=1}^{N}(I_k^2/2)\partial L_k/\partial y_j \tag{14}$$

Obviously, the desired total short-circuit force is furnished by $F_j=(F_{jx}^2+F_{jy}^2)^{1/2}$. Commonly, the geometry of the bus is such that it is symmetrical about the horizontal plane defined by $y=0$. In such case $F_{jy}=0$, and thus $F_j=F_{jx}$.

Derivation of Basic Equations for Individual Subconductors

In practice, the subconductors are usually connected in parallel; in such case the voltage drop on each is the same. The currents in each subconductor can be obtained by use of this fact, as follows.

In an n-phase system comprised of n distinct conductors, one conductor, say a, can be considered as carrying current I_a (the currents are expressed as complex numbers), and the remaining (n-1) conductors can be considered as connected in parallel and carrying the return current. Thus

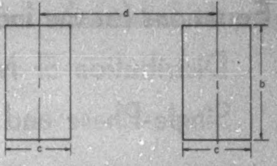


Fig. 2. Cross section of arbitrarily dimensioned rectangular-conductor bus: width of conductor less than height

$$-I_a=(I_b+I_c+\ldots+I_n) \tag{15}$$

Accordingly, as the energy associated with conductor a is given by $W_a=(1/2)\int A w_a dS_a$, the associated inductance, $L_a=$ (reactance drop on conductor a)/($2\pi f I_a$), is

$$L_a=-(2/I_a^2)(w_a^2 S_a^2 \log D_{aa}+ w_a S_a\sum_{i=b}^{n} w_i S_i \log D_a$$

$$=-2\sum_{i=a}^{n}(I_i/I_a)\log D_{ai} \text{ abhenrys per cm of conductor length} \tag{16}$$

The 60-cycle reactance is given by equation 17, as follows

TABLE I

VALUES OF CONSTANTS FOR THE GEOMETRIC MEAN DISTANCE OF A RECTANGLE

b/c or c/b	k	$\log \epsilon$
0.	0.22313	0.
0.025	0.22333	0.00089
0.05	0.22346	0.00146
0.10	0.22360	0.00210
0.15	0.22366	0.00239
0.20	0.22369	0.00249
0.25	0.22369	0.00249
0.30	0.22368	0.00244
0.35	0.22366	0.00236
0.40	0.22364	0.00228
0.45	0.22362	0.00219
0.50	0.22360	0.00211
0.55	0.22358	0.00203
0.60	0.22357	0.00197
0.65	0.22356	0.00192
0.70	0.22355	0.00187
0.75	0.22354	0.00184
0.80	0.22353	0.00181
0.85	0.22353	0.00179
0.90	0.22353	0.00178
0.95	0.223525	0.00177
1.00	0.223525	0.00177

Fig. 3. Values for self geometric mean distances of rectangular areas of Figs. 1 and 2

$$X_a = -240\sum_{i=a}^{n}(I_i/I_a)\log D_{ai}$$

abohms per cm of conductor length

$$= 0.0528\sum_{i=a}^{n}(I_i/I_a)\log_{10}(1/D_{ai})$$

ohms per 1,000 feet of conductor length (17)

If the individual conductors are composite, *a* comprised of *a′* subconductors numbered from 1 to *a′*, *b* of *b′* subconductors numbered from $(a'+1)$ to $(a'+b')$, and correspondingly for the others, equation 16 yields

$$L_a = -(2/I_a^2)\Big(\sum_{i=1}^{a'}\sum_{j=1}^{a'+\ldots+n'} w_i w_j S_i S_j$$

$$\log D_{ij}\Big) \text{ abhenrys per cm of conductor length} \qquad (18)$$

Corresponding expressions for the reactance are obvious from equation 15. The equations of short-circuit force and current associated with each subconductor can be obtained similarly by use of the approximate equations given in the previous section under the title "Derivation of Basic Equations for Individual Conductors."

Dimensional Units

In derivation of the basic equations in the foregoing, convenience stems from taking all quantities in the absolute system: linear dimensions in cm, current in abamperes, inductance in abhenrys per cm of bus length, energy in ergs per cm of length, and force in dynes per cm of conductor length. From equations 2 and 3, however, it is easily established that in calculating inductances the cross-sectional dimensions of the bus and conductors can be expressed in any convenient like unit of length; and that in calculating forces the same can be done for the end force that results as expressed in dynes per unit (chosen for the linear dimensions) of conductor length.

Calculation of Geometric Mean Distances

Grover[3] has advanced tables enabling easy calculation of the self and mutual geometric mean distances of identical parallel-sided rectangles, located as in Figs. 1 and 2. These tables are reproduced in Figs. 3 through 5 for convenience.

The self geometric mean distance *r* is given by

$$\log r = \log(b+c) - 1.5 + \log \epsilon \qquad (19)$$

where $r = k(b+c)$, and $\log k = -1.5 + \log \epsilon$. Values of the parameters *k* and $\log \epsilon$ (here, ϵ is a certain parameter, not the base of natural logarithms) are given in Table I as a function of the ratio b/c or c/b, whichever is less than unity. The corresponding mutual geometric mean distance $R = kd$ is given by

$$\log R = \log d + \log k \qquad (20)$$

FIRST ILLUSTRATIVE EXAMPLE

Calculation of Geometric Mean Distances

In illustration of application of the basic theory derived in the foregoing, consider analysis of the triply subdivided conductor single-phase bus typified in Fig. 6. Each conductor is comprised of three identical subconductors. The dimensions are $b = 1$ cm, $h = 10b = 10$ cm, and $a = 4h = 40$ cm. Use of the tables in Figs. 3 through 5, and interpolating therein by use of forward differences as necessary, yield the 6 self and 24 mutual geometric mean distances. The corresponding numerical values are tabulated in the Appendix.

Calculation of Subconductor Currents

For the bus of Fig. 6, $m = n = 3$ in equation 9. Hence, the inductances associated with the individual subconductors are

$$L_1 = -(2/I_1^2)[I_1^2(\log D_{11} - \log D_{16}) + I_1I_2(\log D_{12} - \log D_{15}) + I_1I_3(\log D_{13} - \log D_{14})] \qquad (21)$$

$$L_2 = -(2/I_2^2)[I_2^2(\log D_{22} - \log D_{25}) + I_1I_2(\log D_{21} - \log D_{26}) + I_2I_3(\log D_{23} - \log D_{24})] \qquad (22)$$

$$L_3 = -(2/I_3^2)[I_3(\log D_{33} - \log D_{34}) + I_2I_3(\log D_{32} - \log D_{35}) + I_1I_3(\log D_{31} - \log D_{36})] \qquad (23)$$

As the subconductors *1*, *2*, and *3* are connected in parallel, the voltage drop on each is the same. Accordingly,

$$(R + jwL_1)I_1 = (R + jwL_2)I_2 \qquad (24)$$

TABLE II

VALUES OF THE GEOMETRIC MEAN DISTANCE OF EQUAL PARALLEL RECTANGLES WITH THEIR LONGER SIDES IN THE SAME STRAIGHT LINE

The values of log *k* in the table are all negative, except in those in the last column, δ=1, (squares).

$\gamma=c/d$	$\delta=0$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	$\delta=1.$
0	0	0	0	0	0	0	0	0	0	0	0
	—	—	—	—	—	—	—	—	—	—	+
0.05	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001	0.0000	0.0000
0.10	0.0008	0.0008	0.0008	0.0008	0.0007	0.0006	0.0005	0.0004	0.0003	0.0002	0.0000
0.15	0.0019	0.0019	0.0018	0.0017	0.0016	0.0014	0.0012	0.0010	0.0006	0.0003	0.0000
0.20	0.0034	0.0033	0.0032	0.0030	0.0028	0.0025	0.0021	0.0017	0.0012	0.0006	0.0000
0.25	0.0053	0.0052	0.0051	0.0048	0.0044	0.0039	0.0034	0.0027	0.0019	0.0010	0.0000
0.30	0.0076	0.0076	0.0073	0.0069	0.0064	0.0057	0.0048	0.0038	0.0027	0.0014	0.0001
0.35	0.0105	0.0104	0.0100	0.0095	0.0087	0.0078	0.0066	0.0052	0.0036	0.0018	0.0001
0.40	0.0138	0.0136	0.0132	0.0125	0.0115	0.0102	0.0086	0.0068	0.0047	0.0024	0.0002
0.45	0.0176	0.0174	0.0169	0.0159	0.0146	0.0130	0.0110	0.0086	0.0059	0.0029	0.0003
0.50	0.0220	0.0217	0.0210	0.0198	0.0182	0.0161	0.0136	0.0106	0.0073	0.0036	0.0005
0.55	0.0269	0.0266	0.0257	0.0243	0.0222	0.0197	0.0164	0.0128	0.0087	0.0042	0.0007
0.60	0.0325	0.0321	0.0310	0.0292	0.0267	0.0235	0.0196	0.0152	0.0103	0.0048	0.0010
0.65	0.0388	0.0383	0.0369	0.0347	0.0316	0.0277	0.0231	0.0178	0.0120	0.0055	0.0014
0.70	0.0458	0.0452	0.0435	0.0408	0.0370	0.0324	0.0269	0.0207	0.0137	0.0062	0.0019
0.75	0.0536	0.0529	0.0509	0.0476	0.0431	0.0375	0.0310	0.0237	0.0156	0.0070	0.0023
0.80	0.0625	0.0616	0.0591	0.0551	0.0497	0.0431	0.0354	0.0269	0.0176	0.0075	0.0031
0.85	0.0725	0.0714	0.0683	0.0634	0.0569	0.0491	0.0401	0.0302	0.0196	0.0081	0.0037
0.90	0.0839	0.0825	0.0786	0.0726	0.0648	0.0555	0.0451	0.0337	0.0216	0.0087	0.0046
0.95	0.0973	0.0954	0.0903	0.0828	0.0734	0.0625	0.0504	0.0374	0.0236	0.0092	0.0056
1.00	0.1137	0.1106	0.1037	0.0942	0.0828	0.0700	0.0561	0.0413	0.0258	0.0098	0.006

Fig. 4. Values relative to rectangular areas located as in Fig. 1

and also the following equation:

$$(R+j\omega L_2)I_2=(R+j\omega L_3)I_3 \qquad (25)$$

Calculating L_1, L_2, and L_3 from equations 21 through 23 by use of the numerical values tabulated in the Appendix; computing from the well-known equation $R=\rho l/A$ the resistance of each copper subconductor as $R=(1.78)10^{-7}$ ohm per cm; and assuming a frequency of $f=50$ cycles (as used in the experimental tests discussed in the following) give

$$(178+j280)I_1-(178+j224)I_2-(j200)I_3=0 \qquad (26)$$

$$(j260)I_1+(178+j284)I_2-(178+j220)I_3=0 \qquad (27)$$

On taking the sum of the subconductor currents I_1, I_2, and I_3 as equal to $100I$ (a convenient normalization), thus

$$I_1+I_2+I_3=100I \qquad (28)$$

the simultaneous solution of equations 26 through 28 gives

$$I_1=(35.11+j5.33)I,\ I_2=(18.75-j14.75)I, I_3=(46.14+j9.42)I \qquad (29)$$

Experimental Confirmation

The bus of the first example is that investigated experimentally by Schwenkhagen[7] in his study of eddy-current losses in subdivided-conductor busses. He found, by measurement, the subconductor currents to be

$$I_1=34.9+j5.2,\ I_2=18.8-j14.6,\ I_3=46.2+j9.4 \qquad (30)$$

Confirmatively, the corresponding calculated values of equation 29 are in excellent agreement with the experimentally determined values of equation 27.

Calculation of Associated Inductances

Substituting the values of the subconductor currents from equation 29 in equations 21 through 23 yields the magnitudes $\bar{L}$ of the inductances L associated with subconductors *1*, *2*, and *3* as

$$\bar{L}_1=(13.15)10^{-9};\quad \bar{L}_2=(20.61)10^{-9};\quad \bar{L}_3=(10.15)10^{-9} \text{ henry per cm of conductor length} \qquad (31)$$

Calculation of Total Inductance

For the bus of Fig. 6, $N=2$, $p_i=3$, and $q_k=3$. By equation 12, the total inductance of each of the two conductors is

$$L=-2\sum_{i=1}^{2}(\pm 1)(1/S_iS_k)\sum_{m=1}^{3}\sum_{n=1}^{3}S_mS_n \log D_{mn} \qquad (32)$$

where (± 1) denotes that for $i=k$, $(+1)$ is to be chosen; and for $i\neq k$, (-1) is to be chosen. Corresponding calculation yields the total inductance of the bus as

$$L=(4.885)10^{-9} \text{ henry per cm of bus length} \qquad (33)$$

Check of Calculated Values of Inductance

As the resistance of each subconductor is small compared with the associated reactance, it is to be expected that a good approximation to the total reactance of the bus can be obtained by ignoring the resistance and computing the approximate reactance as the value of the equivalent reactance of the paralelled subconductors. In such approach the general expression for the correspondingly approximated equivalent total inductance is $L=(L_1L_2L_3)/(L_1L_2+L_2L_3+L_1L_3)$. Now by substituting the complex values pertinent to the numerical magnitudes of equation 31 in this expression and finding the modulus, $\bar{L}=(4.11)10^{-9}$ henry per cm of bus length. Corroboratively, this value is in fair agreement with the accurate value of equation 33; hence, it provides a check both as to the correctness of the computation yielding the value in equation 33 and the aforementioned procedure for approximate calculation of reactance.

A simple procedure which provides an approximate value of the total inductance, and which also enables substantial check of all the foregoing work, is to consider the three subconductors of one conductor as represented by a single conductor of the same cross-sectional area, as indicated in Fig. 7. By equation 9, the total inductance of this "equivalent" bus is

$$L=-(2/I^2)\sum_{i=1}^{2}\sum_{j=1}^{2}I_iI_k \log D_{ij} = (4.5044)10^{-9} \text{ henry per cm of bus length} \qquad (34)$$

This value is in good agreement with the value of equation 33. Often, this

TABLE III

VALUES OF THE GEOMETRIC MEAN DISTANCES OF EQUAL PARALLEL RECTANGLES WITH THEIR LONGER SIDES PERPENDICULAR TO THE LINE JOINING CENTERS

The table gives values of log *k* as a function of $B=b/d$ or $1/\beta=d/b$, whichever is less than unity, and for $\Delta=c/b$.[5] In all cases log *k* is positive.

β	$\Delta=0$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0	0	0	0	0	0	0	0	0	0	0
0.1	0.0008	0.0008	0.0008	0.0008	0.0007	0.0006	0.0005	0.0004	0.0003	0.0002	0.0000
0.2	0.0033	0.0033	0.0032	0.0030	0.0028	0.0025	0.0021	0.0017	0.0012	0.0007	0.0000
0.3	0.0074	0.0073	0.0071	0.0067	0.0062	0.0056	0.0048	0.0038	0.0027	0.0015	0.0001
0.4	0.0129	0.0128	0.0124	0.0118	0.0109	0.0098	0.0084	0.0068	0.0050	0.0027	0.0003
0.5	0.0199	0.0197	0.0191	0.0182	0.0169	0.0152	0.0131	0.0106	0.0077	0.0043	0.0005
0.6	0.0281	0.0278	0.0271	0.0258	0.0240	0.0216	0.0185	0.0152	0.0111	0.0064	0.0011
0.7	0.0374	0.0371	0.0361	0.0344	0.0320	0.0290	0.0251	0.0206	0.0155	0.0090	0.0019
0.8	0.0477	0.0473	0.0461	0.0440	0.0411	0.0373	0.0321	0.0268	0.0200	0.0129	0.0031
0.9	0.0589	0.0584	0.0569	0.0544	0.0506	0.0464	0.0404	0.0338	0.0254	0.0158	0.0046
1.0	0.0708	0.0702	0.0685	0.0655	0.0614	0.0560	0.0492	0.0406	0.0313	0.0199	0.0065
0.9	0.0847	0.0841	0.0821	0.0787	0.0738	0.0675	0.0596	0.0501	0.0382	0.0250	
0.8	0.1031	0.1023	0.0999	0.0959	0.0903	0.0829	0.0745	0.0622	0.0485		
0.7	0.1277	0.1268	0.1240	0.1192	0.1125	0.1037	0.0925	0.0788			
0.6	0.1618	0.1607	0.1573	0.1507	0.1436	0.1329	0.1194				
0.5	0.2107	0.2094	0.2053	0.1984	0.1886	0.1754					
0.4	0.2843	0.2826	0.2776	0.2691	0.2567						
0.3	0.4024	0.4003	0.3942	0.3831							
0.2	0.6132	0.6105	0.6021								
0.11	1.0787	1.1075									
$1/\beta$	$\Delta=0$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.	0.9	1.0

Fig. 5. Values relative to rectangular areas located as in Fig. 2

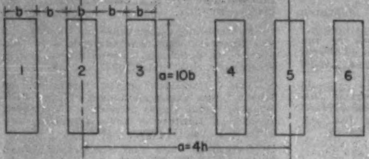


Fig. 6 (left). Cross section of triply subdivided conductor single-phase bus

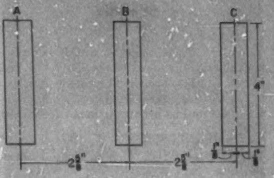


Fig. 8 (right). Cross section of single-conductor 3-phase bus

simplified calculation provides values sufficiently accurate for design purposes.

SECOND ILLUSTRATIVE EXAMPLE

In illustration of analysis pertinent to polyphase busses consider the singly divided 3-phase bus typified in Fig. 8. It is illuminative to consider both negative-sequence and positive-sequence phase currents.

Negative-Sequence Currents

The phase currents are denoted respectively by $I_A=I\epsilon^{j0}$, $I_B=I\epsilon^{-j120}$, and $I_C=I\epsilon^{j120}$. Calculation of the self and mutual geometric mean distances through use of the tables in Figs. 3 through 5 and substitution in equation 16, as indicated, give the inductances associated with each phase as

$$L_A=(2.9162-j1.0200)10^{-9},\quad L_B=(2.3280)10^{-9},\quad L_C=(2.9162-j1.0200)10^{-9} \qquad (35)$$

The resistance of each copper conductor is $R=(525)10^{-9}$ ohm per cm of bus length. The total voltage drop on a closed circuit comprised of the conductors of phase A and phase C is

$$V_{AC}=(R+jwL_A)I_A-(R+jwL_C)I_C =(1{,}931+j1{,}528)(10^{-9})I$$

The average equivalent impedance of phase A to neutral of this closed circuit, as measured from the sending end, is $\bar{Z}_{AC_n}=V_{AC}/\sqrt{3}I=(1{,}115+j882)10^{-9}$, hence $\bar{Z}_{AC_n}=(1{,}420)10^{-9}$ ohms per cm of bus length. Thus, the average equivalent reactance of phase A to neutral is

$$\bar{X}_{AC_n}=(\bar{Z}_{AC_n}{}^2-\bar{R}_{AC}{}^2)^{1/2} =[(1{,}420)^2-(525)^2]^{1/2} =(1{,}320)10^{-9}\text{ abohm per cm of bus length} =0.0402\text{ ohm per 1,000 feet of bus length} \qquad (36)$$

Similar computation yields

$$\bar{X}_{AB_n}=0.0336\text{ ohm per 1,000 feet of bus length} \qquad (37)$$

The voltage drops on the individual conductors are furnished by use of the corresponding associated reactances, to be calculated from $X_n=wL_n$. Hence, equation 35 yields $X_{A_n}=(1{,}100-j384.25)10^{-9}$ $X_{B_n}=(878)10^{-9}$, and $X_{C_n}=(1{,}100+j384.25)10^{-9}$ ohm per cm of bus length. The corresponding magnitudes are

$$\bar{X}_{A_n}=0.0355;\ \bar{X}_{B_n}=0.02676;\ \bar{X}_{C_n}=0.0355\text{ ohm per 1,000 feet of bus length} \qquad (38)$$

Experimental Confirmation

The bus of Fig. 8 is one of several investigated experimentally by Schurig, as remarked in his discussion of reference 3, the pertinent data being advanced in his discussion of Carlson's[4] earlier paper. He obtained a measured reactance of $\bar{X}_{AC_n}=0.041$ ohm per 1,000 feet of bus length. Corroboratively, the corresponding calculated value of equation 36 is in excellent agreement with this measured value.

Positive-Sequence Currents

The phase currents are denoted respectively by $I_A=I\epsilon^{j0}$, $I_B=I\epsilon^{j120}$, and $I_C=I\epsilon^{-j120}$. Analysis similar to that for negative-sequence currents yields

$$V_{AC}=(-410+j2437)(10^{-9})I$$

$$Z_{AC_n}=(1{,}426)10^{-9}\text{ ohm per cm of bus length}$$

$$\bar{X}_{AC_n}=0.0402\text{ ohm per 1,000 feet of bus length}$$

It is of interest to note that the value of $\bar{X}_{AC_n}$ is the same as in equation 36 for negative-sequence currents.

The magnitudes of the associated reactances are

$$\bar{X}_A=0.0355;\ \bar{X}_B=0.02676;\ \text{and}\ \bar{X}_C=0.0355\text{ ohm per 1,000 feet of bus length} \qquad (39)$$

These values for positive-sequence currents are in essential agreement with the corresponding values of equation 38 for negative-phase-sequence currents. This agreement manifests that the magnitudes of the associated reactances of a singly divided 3-phase bus are substantially unaffected by phase sequence.

Epitomization of Third Illustrative Example

Analysis similar to that for the singly divided 3-phase bus of the second illustrative example is detailed in reference 8 for the more complex geometry of the doubly divided 3-phase bus of Fig. 9. Corroboratively, corresponding values of calculated and experimentally determined[9] values of reactance are in excellent agreement.

Conclusions

Because of the substantial agreement of calculated and experimentally determined values in each of the three illustrative examples, it is concluded that derivation of the basic equations has been correctly effected; and that these can be employed in bus design and in the predetermination of bus performance as detailed in the several foregoing examples.

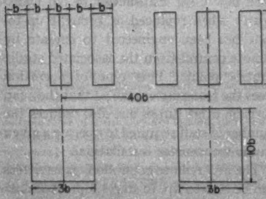


Fig. 7 (left). Approximation of triply subdivided conductor by equivalent rectangular cross section

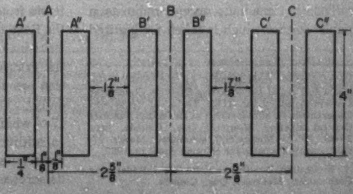


Fig. 9 (right). Cross section of doubly divided conductor 3-phase bus

Appendix

By geometric symmetry, the six self geometric mean distances of the Fig. 6 bus are identical. The numerical value is typified by $\log D_1 = \log(11) - 1.5 + 0.0021 = 0.9$. Values of the mutual geometric mean distances:

$$\log D_{16} = \log D_{61} = 3.7883$$
$$\log D_{15} = \log D_{51} = \log D_{26} = \log D_{62} = 3.7426$$
$$\log D_{14} = \log D_{41} = \log D_{25} = \log D_{52} = \log D_{36} = \log D_{63} = 3.6940$$
$$\log D_{12} = \log D_{21} = \log D_{23} = \log D_{32} = 1.3036$$
$$\log D_{13} = \log D_{31} = 1.6689$$
$$\log D_{24} = \log D_{42} = \log D_{35} = \log D_{53} = 3.6432$$
$$\log D_{34} = \log D_{43} = 3.5897$$

References

1. FORMULAS FOR THE CALCULATION OF THE INDUCTANCE OF LINEAR CONDUCTORS OF STRUCTURAL SHAPE, **Thomas James Higgins.** *AIEE Transactions*, vol. 62, Feb. 1943, pp. 53–58.

2. THE DESIGN OF BUS-BAR INDUSTRIAL DISTRIBUTION SYSTEMS: AN EPITOMIZATION OF AVAILABLE DATA, **Thomas James Higgins.** *Ibid.*, vol. 64, 1945, pp. 385–400; *discussion*, pp. 486–88.

3. EQUATIONS FOR THE INDUCTANCE AND SHORT-CIRCUIT FORCES OF BUSSES COMPRISED OF DOUBLE-CHANNEL CONDUCTORS, **Clifford M. Siegel, Thomas James Higgins.** *Ibid.*, vol. 71, pt. III, Jan. 1952, pp. 522–30; *discussion* by O. R. Schurig, pp. 530–31.

4. *Discussion* by O. R. Schurig of ENCLOSED BUS-BAR ELECTRICAL DISTRIBUTION SYSTEMS FOR INDUSTRIAL PLANTS, **E. T. Carlson.** *Ibid.*, vol. 60, 1941, pp. 700–01.

5. ENGINEERING CALCULATION OF INDUCTANCE AND REACTANCE FOR RECTANGULAR BAR CONDUCTORS, **O. R. Schurig.** *General Electric Review*, Schenectady, N. Y., vol. 36, 1933, pp. 228–31.

6. THE CALCULATION OF THE INDUCTANCE OF SINGLE LAYER COILS AND SPIRALS WOUND WITH WIRE OF LARGE CROSS SECTION, **F. W. Grover.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 17, 1929, pp. 2053–63.

7. UNTERSUCHUNGEN ÜBER STROMVERDRANGUNG IN RECHTECKIGEN QUERSCHITTEN, **H. Schwenkhagen.** *Archiv für Elektrotechnik*, Berlin-Charlottenburg, Germany, vol. 17, 1926–27, pp. 538–89.

8. EQUATIONS FOR THE REACTANCE AND CURRENT DISTRIBUTION OF MULTIPLE-STRAP SINGLE-PHASE AND POLYPHASE BUSSES, **T. H. Chin.** *M.S. Thesis*, University of Wisconsin, Madison, Wis., 1953.

The Canadian Broadcasting Corporation's Radio Service in Canada

N. R. OLDING
NONMEMBER AIEE

THE Canadian Broadcasting Corporation (CBC) is a public corporation established by act of parliament to carry on a national broadcasting service in Canada. The purpose of the CBC is to provide radio and television service for the whole of the nation; to develop Canadian programs and to bring in selected programs from other countries, to make this service available to as many Canadians as can possibly be reached, giving all as uniform a service as possible in English or in French, the two principal languages in Canada. Among the varied broadcasting systems followed by different countries throughout the world, the Canadian system is unique in its use of privately owned and publicly owned radio and television stations as essential parts of an integrated national operation. The continual need to expand the radiobroadcasting coverage in order to provide service to listeners in the remote areas of Canada, to maintain and improve the existing service, and to keep additional expenditures to a minimum has brought about a changing pattern of technical operation which is discussed in this paper.

Today, Canadians accept radio as a public service which is taken for granted along with hydroelectric power and the telephone. In spite of the public interest in television, those engaged in both services know that more Canadians own and operate more radio receivers in 1957 than ever before. Data recently released by the Radio Electronic Television Manufacturer's Association of Canada show that Canadians purchased more than 1,300,000 radio receiving sets in the past two years. The problems brought about by the increase in the number of potential listeners, the continuing need to expand the service to include areas not yet receiving service, and the loss of audience to TV during the evening hours has brought with it the need to operate the CBC's radio facilities on a more economical and efficient basis.

The CBC programs three separate radio networks. Each of these provides service to a group of basic stations and through network extensions a group of supplementary stations. To provide this service the CBC rents, on a contract basis, over 14,000 and occasionally up to 26,000 miles of network facilities.

The Corporation's major network extends from St. John's, Newfoundland, in the East, to Prince Rupert and Vancouver, British Columbia, in the West. For this, the TransCanada network of the Corporation provides 16 hours of program service daily. This network includes 45 stations, 17 of which are CBC owned and operated. In addition to these medium and high powered stations the CBC has extended its service to many remote points by the installation and connection to this network of 52 low-power relay transmitters. The remaining 28 stations which make up the TransCanada group are privately owned and operated. The Dominion Network, which extends from Sydney, Nova Scotia, in the East, to Victoria, British Columbia, in the West, is composed primarily of privately owned stations of which there are 49, and one key station located in Toronto, Ontario, which is owned and operated by the CBC.

The CBC French Network, operating 16 hours daily, is composed of 4 major stations and 2 low-power relay stations owned by the CBC in addition to 20 privately owned stations. This network, extends from Moncton, New Brunswick, in the East to Edmonton, Alberta, in the West. As an additional function and acting for the Government of Canada, the CBC operates two high-power international short-wave transmitters which are located near Sackville, N.B., and are programmed from the Corporation's studios in Montreal. The operation of the four services represents a major engineering function on the part of the CBC, and the two railway communication companies supplying the network facilities.

Technical Operation of Networks

UNATTENDED TRANSMITTER OPERATION

In order to provide as much service as possible with the limited funds available, each new CBC transmitter or old transmitter when replaced or brought up-to-date has been engineered to operate by remote control from the associated studio control point. This type of operation frees the transmitter staff so that they can perform other duties and also reduces the number of staff required to operate a given studio transmitter installation.

The new remote controlled transmitters can be started up, in stages, operated dur-

---

Paper 57-659, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 27, 1957; made available for printing April 18, 1957.

N. R. OLDING is with the Canadian Broadcasting Corporation, Montreal, Que., Canada.

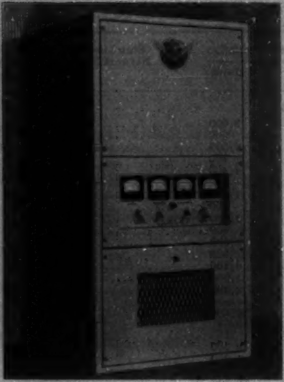


Fig. 1. Canadian Broadcasting Corporation's low-power relay transmitter

ing the broadcast day, and shut down by means of push buttons or switches located in the studio plant. During the operating period, the studio or control technician can readily determine whether or not the readings of several important metered circuits are correct, or if incorrect, are too high or too low. The frequency of the transmitted wave and the transmitted level of the programme can be monitored continuously from the studio control point. If any major unit gives trouble, a visual or audible signal warns the technician on duty at the studio that some fault has occurred.

If anyone attempts to gain illegal entry, if the temperature in the building becomes unusually high, or if fire breaks out, a warning is sounded at the control point, and in some cases a large gong mounted on the outside of the remote transmitter building sounds to scare off prowlers and to attract attention. In the case of fire, a heat sensitive control element closes a circuit which operates motors closing all building air inlets and outlets. When these are closed, carbon dioxide gas (CO_2) is released within the building in sufficient quantity to smother any blaze. Other features include automatic voltage control of voltage sensitive equipment, remote indication showing whether the transmitter tower red warning lights are on or off, and automatic recycling of the transmitter start-up procedure in the event that momentary power surges or lightning discharges knock the transmitter off the air. Manually operated disabling facilities prevent the accidental discharge of CO_2 gas, the ringing of the warning gong or control point signals during periods when authorized personnel are at work in the transmitter building.

CBC stations now operating on a remote control basis include: *CBXA*, Edmonton, Alberta; *CBO*, Ottawa, Ontario; *CBV*, Quebec City, Quebec; *CBAF*, Moncton, New Brunswick; *CBI*, Sydney, Nova Scotia; *CBY*, Corner Brook, *CBT*, Grand Falls, and *CBN*, St. John's, Newfoundland. These normally unattended units are maintained on a prearranged and carefully scheduled basis.

RURAL COVERAGE

On June 1st, 1951, the Royal Commission on the National Development in the Arts, Letters, and Sciences stated: "The Canadian national broadcasting system is the result of ingenious improvision to provide speedily an extensive service in a country where adequate coverage is perhaps the most expensive and the most difficult in the world." Canada is a country of some 16 million people, spread across six time zones. It is bilingual. Almost all of the urban population and the bulk of the rural population live in a narrow belt, about two hundred miles wide but thousands of miles long, across its southern border. As the audience becomes more scattered the cost of providing service to each listener becomes higher because the number of listeners per square mile is reduced. A point is reached where it becomes technically difficult to provide service and the cost per listener becomes exorbitant.

Fig. 2. Canadian Broadcasting Corporation's low-power relay transmitter, rear view

For the most part, the privately owned and operated stations concentrated their

Table I. Rural Versus Urban Coverage

Call	Location	Frequency, Kc	Power, Kw	Daytime Population			Daytime Area, Square Miles		
				Total	Urban	Rural	Total	Urban	Rural
CBU	Vancouver, B.C.	690	10	953,200	771,000	182,200	10,100	430	9,700
CBX	Lacombe, Alta.*	1,010	50	858,000	511,200	346,800	60,900	295	60,600
CBK	Watrous, Sask.†	540	50	1,122,600	381,000	741,600	256,200	420	255,800
CBW	Winnipeg, Man.‡	990	50	787,400	518,400	269,000	53,600	285	53,300
CBL	Toronto, Ont.	740	50	3,680,500	2,749,300	931,200	33,600	810	32,800
CBM	Montreal, Que.§	940	50	4,418,300	3,241,900	1,176,400	52,400	895	51,500
CBF		690	ea.						
CBA	Sackville, N.B.	1,070	50	1,044,600	466,000	578,600	33,600	675	32,900
			310	12,864,600	8,638,800	4,225,800			

NOTE: This list of stations does not represent all of the sound transmitters owned and operated by the CBC. Fourteen other broadcast band transmitters, located at major population centers from St. John's, Newfoundland, to Prince Rupert, British Columbia, 2 short-wave transmitters, 54 low-power relay transmitters, and 4 frequency-modulated transmitters serve to round out the national sound or radio service. Two 50-kw short-wave transmitters operated for the International Service by the CBC provide coverage in several foreign countries. While some of the 50-kw transmitters provide satisfactory service in provinces and time zones outside those in which they are located, there is a practical and economical limit to the extension of coverage by increasing the power output of transmitters.

* Coverage excludes that area which receives a stronger signal from CBK.

† Coverage excludes that area which receives a stronger signal from CBW and CBX.

‡ Coverage excludes that area which receives a better signal from CBK.

§ Coverage quoted is for CBF and is duplicated almost entirely by CBM.

coverage in the larger urban areas, leaving a large and frequently remote rural audience either without service or provided with an erratic and inadequate service. The publicly owned system on the other hand was expected to provide service in most of these rural areas as well as in the major urban centers. Fortunately, the CBC was able to provide a large percentage of listeners with service by strategically located 10-kw to 50-kw transmitters. The locating of these transmitters presented many problems, and the final location selected for the transmitter site was always a compromise between the best rural coverage and satisfactory urban coverage. The rural versus urban coverage so provided can best be illustrated by some typical examples as in Table I.

Low-Power Relay Transmitter Operation

Field tests made with a low-power transmitter set up in a mountain valley in British Columbia in the summer of 1940 confirmed a CBC engineer's belief that a small transmitter with a power output of 20 to 25 watts could provide a reasonably satisfactory service in isolated communities where the more concentrated population did not exceed 10,000.

While the normal range of a station of such limited power would not exceed a radius of 3 to 5 miles, it was found that a small part of the radiated power could be introduced into the local telephone, telegraph, and carrier circuits by locating the transmitting antenna within 50 feet of the overhead wires and that this could be done without creating serious interference with these services. By a further utilization of the carrying effect of the open copper-wire lines, it was possible to provide usable service to small communities, ranchers, farmers, sawmill workers, and railroad men where the residences were located near the right of way of such lines and up to a distance of 60 miles from the low power transmitter site. By this means, many previously unserved areas, through which the network lines passed or which could be tied in to the network by means of a stub line, could be provided with network program service.

The first station of this type to be placed in operation was installed in Revelstoke, B.C., on October 22, 1940. This initial unit had not been designed for 16 hour a day, unattended service. In view of these and other shortcomings, and as many small communities were demanding service, a transmitter unit was designed especially for the purpose. Some of the requirements were as follows:

Table II. Low-Power Relay Transmitters

Call	Location	Approximate Population		Stub Line from Nearest Network Repeater Station, Miles
		Urban	Rural	
		Ontario*		
CBLJ	Jamestown		3,000	65
CBLG	Geraldtown	3,600	900	46
CBLR	Red Rock	1,400	2,600	47
CBLC	Chapleau	2,700	700	
CBLE	Beardmore		1,400	48
CBLN	Nakina		800	28
CBLS	Sioux Lookout	2,700	1,300	46
CBLH	Hornepayne	1,300	200	50
CBLL	Longlac		1,500	18
CBLW	White River		700	46
CBLB	Schreiber	2,500	1,500	41
CBLF	Foleyet		900	
CBLA	Atikokan		3,100	111
CBLD	Dryden	2,900	1,900	90
CBLM	Marathon	1,500	2,300	51
TOTAL	15	18,600	22,800	
		New Brunswick		
CBAB	Grand Falls	2,700	9,200	
CBAM	Edmunston	12,000	1,000	
TOTAL	2	14,700	10,200	
		Quebec		
CBFG	Gaspe	1,900	13,700	
CBFB	Megantic	6,800	6,800	
TOTAL	2	8,700	20,500	
		Alberta		
CBXB	Banff	5,100	400	
CBXL	Blairmore	5,100	3,600	
CBXJ	Jasper	2,300	100	
CBXC	Coleman	2,600	2,800	
TOTAL	4	15,100	6,900	
		British Columbia†		
CBUG	Kaslo		1,200	
CBRF	Fernie	2,500	3,000	
CBRO	Greenwood	1,900	5,000	
CBXE	Golden		1,500	
CBXM	McBride		1,700	
CBUC	Salmon Arm	2,600	12,300	
CBRM	Creston	1,800	4,800	
CBUE	Hope	1,900	3,600	
CBUK	Kitimat		8,000	
CBRN	North Bend		1,300	
CBRQ	Quesnel	1,800	3,700	
CBRA	Revelstoke	3,200	800	
CBRJ	Grand Forks	1,900	4,500	
CBRL	Williams Lake		2,100	
CBRR	Cranbrook	4,100	2,700	
CBUF	Ocean Falls	‡3,200	100	
CBRG	Prince George	§5,300	3,200	
CBRK	Kimberley	6,600	2,700	
CBUB	Osoyoos		2,600	
CBRE	Lytton		700	
CBUD	Castlegar	1,500	5,300	
CBUA	Oliver	1,100	5,800	
CBXN	Natal		2,600	
CBRS	Smithers	1,300	1,100	
CBRV	Vanderhoof		2,300	
CBRB	Burns Lake		2,300	
CBRC	Terrace		2,100	
CBRD	Field		700	
CBRP	Princeton	2,400	2,300	
CBRT	Ashcroft		1,700	
CBRH	Hazelton		600	
TOTAL	31	43,100	92,300	

NOTE: Population figures were obtained from the audience research department of the CBC and are based on the Canadian Census of June 1st, 1951 projected to June 1st, 1955 or census figures for 1956. The 1951 Census definition of urban was used to determine the urban population. This definition read as follows: "The urban population in 1951 includes all persons residing in cities, towns and villages of 1,000 and over, whether incorporated or unincorporated, as well as the population of all parts of census metropolitan areas."

* Total watts output 300. Potential audience per watt of output, urban 62 and rural 71.

† Total watts output 620. Potential audience per watt of output, urban 70 and rural 150.

‡ Network unavailable served by S.W. and Broadcast Band from Vancouver. Frequency diversity receiver.

§ This town now has a small privately owned broadcast transmitter in addition to the CBC relay unit.

1. Since the contract for the national radio network service had been placed with the two major Canadian railway companies, the transmitter would have to feed an antenna which could be conveniently and unobtrusively erected on railway station property

rather than one designed for maximum efficiency.

2. The antenna tuning and coupling unit should provide a direct copper-circuit path from antenna to ground so that static charges could not build up and break down components.

3. The power output should be between 15 and 25 watts and antenna coupling facilities should permit this output to be fed to antennas having a wide range of reactance and resistance values.

4. Some form of lightning gap and choke should form a part of the installation.

5. The tube line-up should be as simple and efficient as possible.

6. The unit should provide long and reliable service when operated on poorly regulated or fluctuating sources of power supply.

7. A minimum of meters and panel controlled adjustments, consistent with reliable operation, should be used.

8. The power supply should be broken down into fused and sectionalized units to provide good regulation and ease of servicing.

9. The mechanical construction and wiring harness should permit sectionalized chassis units to be removed and replaced readily.

10. The audio frequency response should be equal to, or better than, that of the associated network lines, that is, ±2 decibels from 100 to 5,000 cycles.

11. The audio input should be arranged to provide bridging input or direct termination across a 600-ohm line and the built-in speech input amplifier should permit operation at 90% modulation with line levels as low as −10 to −20 decibels.

12. All components should be operated well below rated capacity.

13. Servicing of the unit should be as simple as possible.

14. The unit should be capable of operating for periods of 4 months or longer between periodic maintenance inspection calls.

As no ready-made equipment incorporating these features was available on the market, arrangements were made to co-operate with a radio manufacturer in the design and construction of a prototype unit. When this proved satisfactory, additional units were manufactured and a detailed "Maintenance Manual" was prepared so that technicians, with little or no prior experience with this type of equipment, could locate and repair minor faults and also could replace defective tubes and fuses.

The first of the new units was installed in Revelstoke in July 1942, and a unit of this type has been in daily operation since that date. This system was pioneered in Canada and 54 units are now in operation providing national network program service for 16 hours a day to some 62,500 additional radio homes and a potential audience of approximately 250,000 listeners; see Figs. 1 and 2, and Table II.

Many valuable suggestions have been received from technicians and engineers in the field whose duty it is to install and maintain these units. From time to time improvements have been made in the basic design with the result that new or converted units will normally provide uninterrupted service, other than failure of primary power service, for periods of 8 months to well over a year. In practically every instance, these relay installations have exceeded expectations, providing program service in many towns and villages located 25 to 35 miles from the transmitter location.

While these miniature transmitters have been designed to operate with little or no attention, it has been CBC practice to install them in telephone, telegraph, or railway stations, and arrangements have been made with the technical or nontechnical staff at each point to provide the minimum of service required to assure satisfactory operation. In the event of unexpected failure of tubes or other components, arrangements have been made, at most points, whereby a qualified local technician or citizen will provide emergency repair service. If the local technician cannot locate or repair the fault, a telegram is sent to the nearest CBC supervisory office and a technician is dispatched immediately to locate and repair the fault.

As a matter of interest, the high-power transmitters previously listed provide service to 41.5 potential urban listeners and 13.6 rural listeners per watt of output power. The annual operating and maintenance cost excluding network costs and depreciation but including salaries of staff, maintenance of buildings, grounds, tube and part replacements, power, heating, etc., averages about $7.70 a year per 100 listeners.

The low-power relay transmitters in Ontario operating to an audience prepared to accept a somewhat less exacting service provide service to an urban audience of approximately 62 and a rural audience of 71 per watt of output. The average operating and maintenance cost, excluding network and depreciation costs, is approximately $3.04 a year per 100 potential listeners.

When stub network costs, from the closest contact with the TransCanada network are added to the operating cost, the cost of serving each potential listener in the typical group served by the 15 relay transmitters in Ontario, is approximately $1.85 per year. The average cost of the power, space, limited service, semiannual inspections, emergency trips, tube, and equipment replacements and other costs excluding depreciation, averages, in the case of the above 15 units, $1,254.40 per unit per year of which sum $54.40 covers the cost of all inspections, maintenance, and emergency trips, tube and component replacements, and local emergency service.

Conclusions

The data presented in this paper and the experience gained in the operation of the equipment described lead to the following conclusions.

1. Radiobroadcast transmitters, when designed for operation by remote control and when housed in a building designed to permit this type of operation, can provide service equal in all essential respects to similar attended units.

2. 20-watt unattended or semiattended relay transmitters can provide a useful radiobroadcast service in areas remote from the larger centers of population and will, with very few exceptions, provide this service for periods of 4 months or more between periodic maintenance inspections.

3. Either or both of the above forms of operation can reduce technical manpower requirements and operating costs.

Reference

1. BRITISH COLUMBIA'S BROADCAST RELAY SYSTEM, N. R. Olding. *Electronics*, New York, N. Y., Sept., 1944, pp. 92–97, 304.

The D-C Controlled A-C Voltage Source, A New Magnetic Amplifier

E. W. MANTEUFFEL
ASSOCIATE MEMBER AIEE

R. O. McCARY
ASSOCIATE MEMBER AIEE

MAGNETIC AMPLIFIERS have become increasingly popular in the last few years, and extensive work has been done to relate their performance to various operating conditions. There has, however, been a distinct tendency to confine the work to fairly well-known basic circuits. These have been so exhaustively studied that there appears to be little hope of greatly improving their operating characteristics by various modifications.

This paper describes a new type of magnetic amplifier, exhibiting many operating characteristics unlike those of conventional magnetic amplifiers. These differences may be traced to the inherently different modes of operation: i.e., a conventional magnetic amplifier acts as a nonlinear and controllable impedance or a controllable current gate, while the amplifier to be described acts as a controllable voltage source. It is therefore inherently better suited for applications involving precision voltage regulation such as regulated d-c power supplies.

Method of Operation

The operating principle of the amplifier stems from the magnetic voltage stabilizer whose circuit is shown in Fig. 1. This stabilizer consists of a nonlinear oscillating circuit, capacitor C in parallel with saturable reactor S, and is supplied from a source in series with a linear inductance L. The load resistor R is connected in parallel to the ferroresonant circuit. Since this magnetic voltage stabilizer has been treated elsewhere,[1] it will not be described in detail here; if the hysteresis loop of the core material of S is quite rectangular, the half-wave average output voltage E_0 is inherently constant over a wide range of input voltages, provided the supply frequency is constant. This voltage is given by

$$E_0 = 4fNAB_s 10^{-8} \tag{1}$$

where

E_0 = volts, average
f = frequency, cycles per second
N = turns
A = effective core cross section, square centimeters
B_s = saturation flux density, gauss

The approximately square wave shape of this voltage across the saturable reactor is illustrated in Fig. 2. At the end of each half-cycle a pulse discharge current I_p flows through the capacitor and saturated reactor.

It is apparent that if the total flux swing in the core could be controlled, a variable voltage source would be available, which would principally retain the inherent advantages of a magnetic stabilizer. A circuit allowing this control is indicated in Fig. 3. The difference between the circuits of Figs. 1 and 3 is the replacement of the saturable reactor S by two saturable reactors S_1 and S_2 with their associated rectifiers CR_1 and CR_2. If no control current flows through the control winding, the total flux swing each half-cycle will be between only positive residual and positive saturation. Only a small voltage will, therefore, appear across the output terminals to which the load R is connected. Actually, this voltage will depend primarily on the saturated inductance of the cores and the resistance of the windings, and rectifiers.

If an appropriate control current is established in the control winding in order to reset the flux during the nonconducting half-cycle, a flux change will occur in saturable reactors S_1 and S_2. A larger voltage than occurs with zero control current will now appear at the output terminals. If the control current is made large enough to reset the cores to negative saturation, the output voltage will reach its maximum value. This voltage is then the same as that of the simple voltage stabilizer as defined by equation 1. The transfer characteristic of the amplifier is illustrated in Fig. 4. Unlike conventional amplifiers, it is apparent that minimum output exists at zero control current.

Since saturation of either core, similar to a magnetic stabilizer, will always occur when the capacitor discharges at the very end of its respective gating half-cycle, voltage must be sustained across its gate winding practically during the full half-cycle. This is not true for conventional magnetic amplifiers where the major flux change takes place during the prefiring interval. While regulation in known magnetic amplifiers is achieved by varying the firing angle, similar to a thyratron, the output voltage of this amplifier is changed predominantly by variation of its amplitude.

The flux resetting operation of conventional magnetic amplifiers of the full-wave self-saturating type depends greatly on the relative value of control circuit impedance. If the control circuit is of low impedance, resetting of one core occurs predominantly during the prefiring interval of the other core. Otherwise, at high control circuit impedance, resetting is done practically throughout the full nonconducting half-cycle.

In contrast to conventional magnetic amplifiers, resetting of flux within the new magnetic amplifier is only slightly influenced by the value of control circuit impedance. Whether this impedance is large or small, voltage across the gate windings is always sustained throughout a complete half-cycle. Therefore, resetting will also take place during full half-cycles. This may be understood also from the fact that the firing angle must always be considered to be close to 180 degrees. Fig. 5 illustrates voltage and current relationships for the new amplifier. The terminal voltages E_{s1} E_{s2} of the two gate windings are at anytime about equal to the capacitor voltage E_c except for minor voltage drops caused by the magnetizing and pulse currents within the rectifiers CR_1 and CR_2. These currents, Is_1 and Is_2, and their components Im_1, Im_2, and Ip_1 and Ip_2 are shown in Fig. 5. The total current I_s is furnished from the capacitor.

In the case of low control circuit impedance, where both gate windings are tightly coupled by the control winding, it is easy to understand that the voltages Es_1 and Es_2 must differ by small voltage drops only. Otherwise, if the control circuit impedance is high, recourse may be taken to the widely used technique of following through a cycle knowing that for steady state the same conditions must exist at the beginning and at the end of the cycle.

At $t=t_0$ in Fig. 5, all the energy of the oscillating circuit is stored in the magnetic field of reactor S_2, since the voltage across the capacitor is zero and the current Ip_2 through S_2 is at its maximum. As E_c builds up from t_0 to t_1, a small magnetizing current, Im_1 flows through reactor S_1

Paper 57-768, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 25, 1957; made available for printing June 11, 1957.

E. W. MANTEUFFEL and R. O. MCCARY are with the General Electric Company, Ithaca, N. Y.

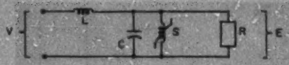


Fig. 1 (left). Magnetic voltage stabilizer

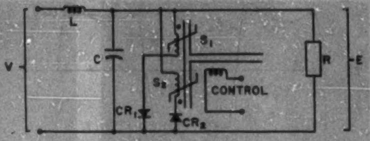


Fig. 3 (right). D-c controlled a-c voltage source

changing its flux from its preset level to saturation. Neglecting the small voltage drop across the rectifier, CR_1, E_{s1} is equal to E_c during this interval. At t_1 the impedance of S_1 drops to approximately that of an air-filled coil and capacitor C begins a damped oscillatory discharge through this small inductance. Since pulse current Ip_1 flows through CR_1 from t_1 to t_2, during this period E_{s1} is still equal to E_c, again neglecting the small voltage drop within the rectifier. At t_2, rectifier CR_1 blocks and flux resetting starts within core S_1, while the flux in S_2 is now building up as rectifier CR_2 becomes conducting.

If, between t_2 and t_3, the rate of change of flux within S_1 would be at any time greater than in S_2, the voltage E_{s1} would attempt to exceed E_c and the rectifier CR_1 would in turn again become conducting. Therefore, E_{s1} may not at any time become larger than E_c. Alternatively, if the average value of E_{s1} between t_2 and t_3 would be less than E_c, the flux level in the core at t_3 would be different from that at t_0. This is not allowable under the assumption of steady state operation.

In the preceding paragraphs it was shown that E_{s1} and, similarly, E_{s2} are the same as E_c throughout the cycle, at least to a close approximation. An observation with an oscilloscope of the voltage difference between E_c and E_{s1} across CR_1 shows only a small drop when the pulse current occurs, and a narrow spike of reverse voltage of about 15 to 20% of the maximum value of E_c. This spike occurs immediately upon cessation of the pulse current. Since reverse voltage appearing across the rectifiers is little and also limited in time to only a small fraction of the cycle, the transfer characteristic of this amplifier is nearly independent of variations in reverse resistance. This is true even for the case of selenium rectifiers.

Design Considerations

To enter into a detailed discussion of design criteria is beyond the scope of this paper, but it is desirable to indicate some of the major considerations. The maximum voltage has been indicated in equation 1; the minimum (at zero control current) is normally about 10 to 15% of maximum. The control winding is chosen in the same way as for a conventional magnetic amplifier.

The input inductor and capacitor must be carefully selected to provide ample output power without unnecessarily increasing losses. The input inductor L must be sufficiently large to provide the required isolation between the source and the oscillating circuit, yet small enough to permit transfer of the required energy from the source to the load. This requirement is met if

$$0.5 \le \frac{WL}{R} \le 0.7 \qquad (2)$$

Values somewhat outside this range may be used, but these values represent normal conditions. Similarly, the capacitor C must be large enough to handle the required power, but not so large as to increase inordinately the losses of the amplifier. A normal range of values is approximately

$$0.5 \le W^2LC \le 2.3 \qquad (3)$$

Again, values somewhat outside this range may be used for some special purpose. It is, however, extremely undesirable to approach closely the condition $W^2LC=1$, since self-quenched oscillations will result from this critical condition.

In order to select an appropriate wire size, it is necessary to determine the rms (root mean square) current flowing in the gate windings of the saturable reactors. Consider the voltage E_{s1} in Fig. 5. Between t_0 and t_1, the flux has been changing in the core. The only current during this time is the magnetizing current I_m. This current may be neglected since it is very small. At t_1, the voltage E_c across the capacitor is a little larger than, but may be approximated by, the average a-c voltage E_0. Upon saturation, the impedance of S_1 becomes much less than that of L or R, and the latter two may be assumed infinite by comparison (implying negligible energy interchange between source and oscillating circuit or load and oscillating circuit) during the period of saturation of S_1. This condition is represented by the circuit of Fig. 6, where E_0 exists as a charge on the capacitor C, and the switch SW closes upon saturation (at t_1 in Fig. 5). The inductor L_0 and the resistor r_s represent the saturated inductance and the loss resistance of the satu-

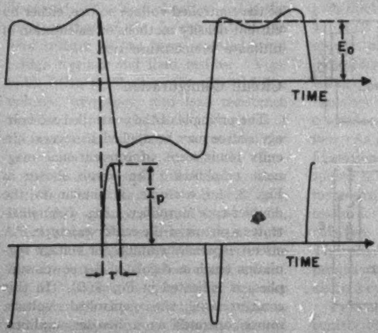


Fig. 2 (left). Current and voltage for saturating core of Fig. 1

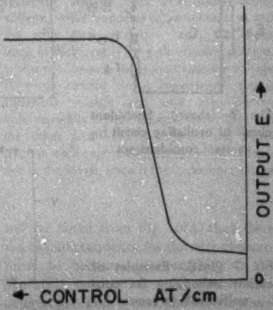
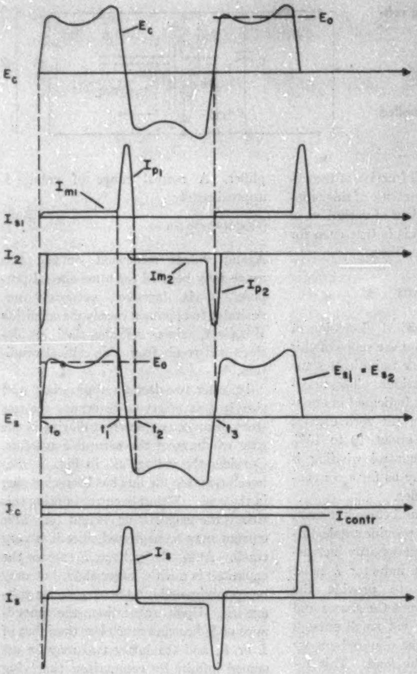


Fig. 4 (right). Generalized transfer characteristic of controlled magnetic-voltage source



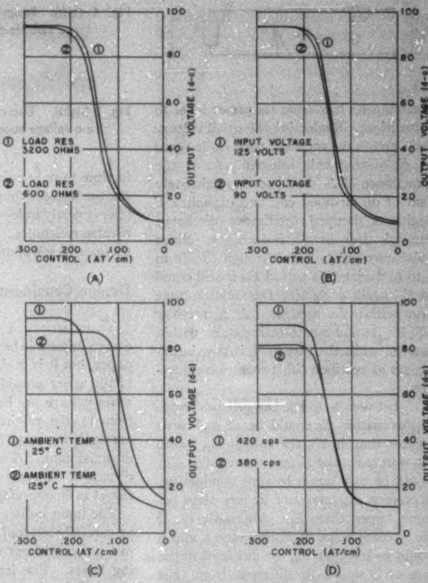


Fig. 5 (left). Voltage and current relationships

Fig. 8 (above). Effect of variations on amplifier characteristic

rable reactor S_1 and its associated rectifier CR_1. If the resistor r_s is assumed to be small, the first current peak of the oscillatory discharge may be approximated by

$$I_p \cong E_o\sqrt{C/L_o} \qquad (4)$$

However, this oscillation will persist only for one half-cycle of its natural frequency (from t_1 to t_2), because the rectifier blocks the subsequent half-cycle. The duration of the current pulse (from t_1 to t_2) is a half period of the natural frequency, i.e.,

$$t_2 - t_1 = \pi\sqrt{CL_o} \qquad (5)$$

Using equations 4 and 5, and remembering that only one current pulse occurs in S_1 every full cycle of the power frequency f, the rms current in S_1 is

$$I_{s_1}(\text{rms}) = \left[\frac{I_p^2}{2}\pi\sqrt{CL_o}\,f\right]^{1/2} = \left[\frac{\pi}{2}E_o^2\frac{C^{3/2}}{L_o^{1/2}}f\right] \qquad (6)$$

This current may be used to determine the necessary wire size for the gate windings of the controlled voltage source, either by current density methods or calculation of ultimate temperature rise.

Circuit Configuration

The principle of the controlled a-c voltage source may be applied to several circuits reminiscent of conventional magnetic amplifiers. The circuit shown in Fig. 3, for example, is similar to the doubler-type amplifier. Fig. 7(A) illustrates a circuit of the center-tap type. A circuit especially suitable for voltage regulation (such as regulated d-c power supplies) is indicated in Fig. 7(B). In this configuration, the controlled voltage source operates as a booster-amplifier,

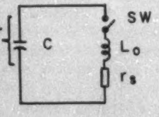


Fig. 6 (above). Equivalent circuit of oscillating circuit for transient considerations

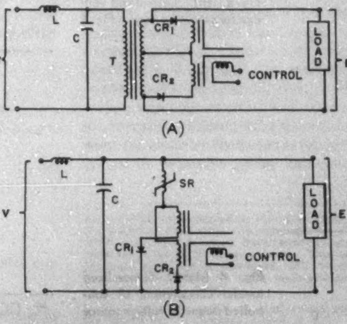


Fig. 7 (right). Examples of circuits embodying the controlled-voltage source

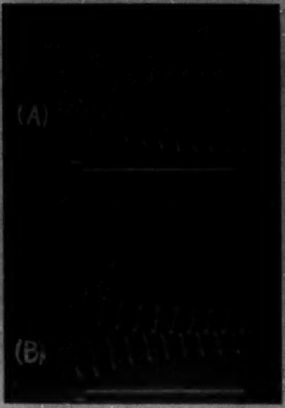


Fig. 9. Transient response oscillograms. Solid line shows control signal

A—Increasing output voltage
B—Decreasing output voltage

since its controlled output voltage is added directly to the fixed voltage appearing at the terminals of a saturable reactor *SR*. The output voltage *E* is controllable over a range determined by the controlled voltage source, but with a minimum value given by the saturable reactor *SR*. Such a design reduces the required power handling capability of the control component, allowing the use of a smaller amplifier, and results in improved time response.

Experimental Evaluation

The controlled voltage source has been tested in several different circuits. The results from two of these circuits will be discussed. The first circuit tested is the same as that of Fig. 3, except that the load resistor *R* has been replaced by a bridge rectifier and load resistor. Figs. 8(A), 8(B), 8(D) show the effect of line voltage, frequency, and load resistance variation. Since the basic advantages of the magnetic stabilizer are retained for this type of amplifier, its transfer characteristic is very little influenced by large variations of either supply voltage or load. Frequency changes, within usual limits, affect mainly the maximum output voltage but are negligible with respect to the linear portion of the characteristic. Fig. 8(C) shows changes in the transfer characteristic expected with variations in ambient temperature. The core material used in these cases is 50% Ni-Fe, grain oriented toroidally wound tape of 0.002-inch thickness.

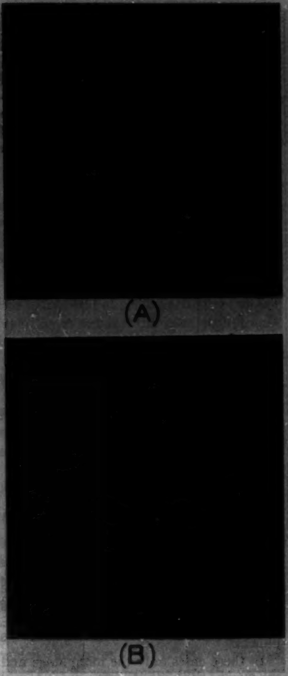


Fig. 10. Output waveform

A—Shows output, one-half, and one-third
B—Shows full output, three-fourths, one-half, and one-fourth, for amplifier with high forward resistance switching rectifiers

Fig. 9 illustrates typical transient responses of this amplifier, the solid trace indicating the time of imposition of a voltage step in the control circuit. The transient response of this amplifier may be computed as that of a conventional amplifier. As expected with the rectifier configuration involved, the decreasing response is somewhat more rapid than the increasing response. Oscillograms were taken for variations in output voltage between 45% and 90% of full output. The power gain was determined to approximately 400 and the time constant is about 2 cycles or 5 milliseconds.

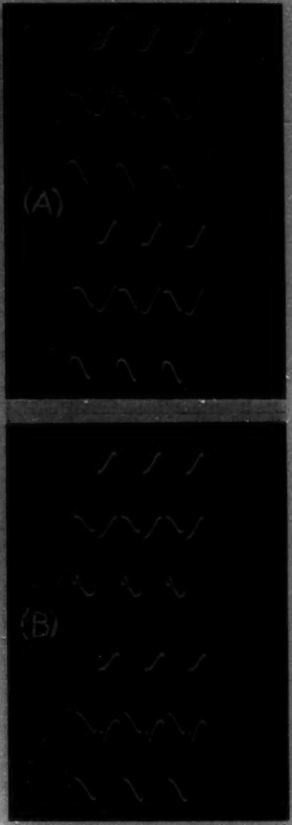


Fig. 11. Capacitor voltage and gate-winding voltage

A—Low-control circuit impedance: top is capacitor voltage and one gate-winding voltage, while bottom is capacitor voltage and the other gate-winding voltage; the upper trace in each case is the gate-winding voltage, while the lower trace is the capacitor voltage
B—High control circuit impedance: top is capacitor voltage and one gate-winding voltage, while bottom is capacitor voltage and the other gate-winding voltage; the upper trace in each case is the gate-winding voltage, while the lower trace is the capacitor voltage

Typical output wave shapes at various control levels are shown in Fig. 10. It will be noted from Fig. 10(A) that there is a definite tendency for the output waveform to become asymmetrical as the operating point travels down the characteristic. This does not affect the trans-

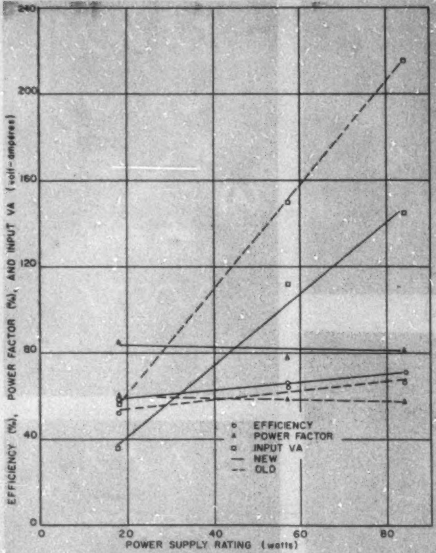


Fig. 12. Comparison of efficiency, power factor, and input volt-ampere requirements of equivalent controlled voltage source amplifiers and standard magnetic amplifiers

fer characteristic, provided the half-wave average output voltage is measured or the output is rectified. This phenomenon has been discussed and has been theoretically studied for the case of the basic magnetic stabilizer by Buchhold and Stuhlinger.[1] If, however, it is necessary to provide a more symmetrical waveform at the lower end of the transfer characteristics this can be achieved by introducing small resistors in series with the switching rectifiers. Oscillograms of Fig. 10(B) were obtained under this condition. A lighter load will also improve wave shape symmetry.

Fig. 11 shows oscillograms of the capacitor voltage and the voltage across both gate windings. Oscillograms were taken at about 50% of full output. Fig. 11(A) represents the case of very-low-control circuit impedance while Fig. 11(B) was taken at high impedance. In both cases, reset occurs during the full nonconducting half-cycle.

The second circuit tested is similar to Fig. 7(B). Several magnetically regulated direct current power supplies have been developed utilizing this circuit configuration with appropriate additions of an output rectifier and filter. Some of these replaced magnetically regulated supplies which utilized conventional magnetic amplifiers.

In these cases only the main stage and the filter were changed. The filter change was made possible by the improved voltage wave-shape, which permits some reduction (about 40%) in required ripple attenuation. Comparative performance measurements on these two sets of magnetically regulated power supplies culminated in these interesting results:

1. The transient response of the new-type supplies is faster by 20% to 50%. (This is primarily because of the reduced lag in the filter, since it has less total energy storage.)

2. The transient response is essentially independent of the level of input voltage and load. This is not true for the standard amplifier supplies

3. For step changes in input voltage, the new-type supplies have a maximum output deviation which is smaller by a factor of about three.

4. The variation of output with frequency (steady state) in the new supplies is about twice that of the old supplies. This is caused by the use of a main stabilizer, *SR* in Fig. 7(B), which forces the booster amplifier to change its operating point.

5. The variation of output with line voltage (steady state) in the new supplies is one-half to one-third that of the old supplies.

6. The variation of output with load (steady state) in the new supplies is about one-fourth that of the old supplies.

7. Power factor and efficiency are also improved in the new supplies, as may be seen from Fig. 12.

Conclusions

The d-c controlled a-c voltage source constitutes a new approach to the improvement of magnetic amplifier performance. It provides an inherently different type of operation as compared to magnetic amplifiers previously known. Advantages of this new amplifier are:

1. There is greater stability of output with variations, transient or steady state, of either line voltage or load.

2. Input requirements are reduced, as a result of improved power factor and efficiency.

3. Filter requirements, where a d-c output is required, are reduced.

4. Often reduced gain may be used, for there is greater inherent stability as mentioned in 1.

5. Influences of current wave shape on source voltage wave shape are reduced.

6. The input inductor has a fail-safe feature that limits input current in case of short-circuited output. Input current is normally limited to perhaps two to three times normal. By proper design it can be held to a lower value.

The foregoing list indicates particular applicability as a main stage in a control or regulating system.

Reference

1. OSCILLATING CIRCUIT INCORPORATING A CHOKE WITH RECTANGULAR MAGNETIZATION CURVE, T. Buchhold, E. Stuhlinger. *Technical Report No. 70*, Redstone Arsenal, Huntsville, Ala., Jan. 15, 1952, pp. 1–44.

A Self-Balancing Transformer-Core-Loss Bridge

A. H. FOLEY
ASSOCIATE MEMBER AIEE

Synopsis: Measuring transformer core loss by impedance bridge methods offers a number of advantages over conventional wattmeter methods, particularly in automation. Human errors are minimized and test accuracy is improved. The bridge is self-balancing and control can be effected either by manual or punched card programming. Direct readings, high precision, and high output torque make the self-balancing bridge particularly suitable for automation.

WHEN power-loss measurements are to be made on electric apparatus, some form of wattmeter is almost invariably chosen. That impedance bridges often may be used with advantage for this purpose is not so well appreciated.

Where automation must be considered, it generally is necessary to look for some form of instrument other than the indicating wattmeter for here no operator will be present to take readings. Further, the instrument may be called upon to perform such tasks as recording data, sensing limits, and others requiring mechanical energy.

The self-balancing impedance bridge has a number of advantages over other forms of instrumentations in applications of this kind. Among these advantages are high mechanical output torque, high precision, reduced sensitivity to test voltage variations, and extreme flexibility of control for setup of ranges.

This paper is primarily concerned with the application of impedance bridges to the measurement of core loss in transformers operated principally at 60 cps (cycles per second). Application to measurements at other frequencies is readily accomplished. Consideration is given toward the design and operation of a self-balancing bridge suitable for use in automation.

Considerations in Applying Bridges to Transformer-Core-Loss Measurements

Before discussing any form of bridge circuit for the measurement of transformer core loss, two points about impedance bridges must be established.

1. Impedance bridges do not measure power loss directly. In general, resistance and reactance are measured and energy loss must be computed.
2. Impedances are measured at fundamental frequency only.

Relative to the first point, where energy loss is specified at some particular voltage level, as is usually the case with transformer core loss, the relation

$$\text{Watt loss} = P = \frac{E^2}{R_p} \tag{1}$$

is convenient. R_p in this case is the effective parallel resistance in the impedance under test and E is the applied voltage.

Where the loss is specified at some current level, the relation

$$P = I^2 R_S \tag{2}$$

is more convenient. R_S is the effective series resistance of the impedance under test and I is the current at which loss is specified.

For core-loss measurements it is desirable, therefore, to choose a bridge network which measures the effective parallel resistance as seen looking into the excitation terminals of the transformer under test. A network of this type is described in this paper. Also, it is shown how the bridge can be made to perform the power loss computation and give direct readings of the desired results.

Relative to point 2, with most impedance bridges, a balance, once it has been made, is valid only at the frequency at which it was made. This generally is true because of the frequency-sensitive nature of some of the parameters in either the bridge network, the impedance under test, or both. Almost invariably waveform distortions exist to some degree in either the current or voltage waves, or both, associated with the impedance under test. This is particularly true if the impedance under test is nonlinear even though the voltage impressed across the bridge is sinusoidal. When a so-called balance has been made a residue of harmonic voltages appear at the bridge detector terminals. To permit the balance to be made with sufficient precision, the bridge balance detector system must have a suitable degree of discrimination against these harmonic voltages. It is because of these mechanisms that impedance bridges measure impedances apparent at fundamental frequency only.

DISTORTED WAVEFORMS

Consider the case where both the current through and the voltage across the impedance under test contain harmonics of the fundamental. Power dissipation is given by the equation

$$P = \sum_{m=1}^{m} E_m I_m \cos\theta_m \tag{3}$$

where P is the total energy dissipation in watts and the subscripts refer to the harmonic order in each term of the summation.

If, however, a sinusoidal voltage at fundamental frequency is impressed across a nonlinear impedance, only the current wave will contain harmonics of the fundamental. With no harmonic voltages present other than the fundamental, equation 3 reduces to the following equation

$$P = E_1 I_1 \cos\theta_1 \tag{4}$$

where the subscript refers to fundamental frequency only. The harmonic currents contribute nothing to the power loss in this case.

A more complete discussion of the mechanisms of harmonic losses in circuits containing nonlinear impedances is given in the references.[1–3]

From the foregoing discussion it is clear that, if energy losses in nonlinear impedance are to be determined with negligible errors by bridge methods without having to apply difficult waveform corrections, the applied voltage must essentially be sinusoidal.[1] A further discussion of the effect of variations of waveform is given in the section on "Bridge Characteristics and Performance."

For transformer-core-loss measurement, the requirement of good voltage waveform does not necessarily place the impedance bridge at a disadvantage. Transformer core loss generally is specified in terms of sinusoidal excitation. Though various methods have been described for correcting loss measurements to a sine-wave basis where distortion is present, sinusoidal excitation still is to be preferred for good accuracy.[4]

Paper 57-559, recommended by the AIEE Indicating and Integrating Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 8, 1957; made available for printing May 31, 1957.

A. H. FOLEY is with the General Electric Company, Pittsfield, Mass.

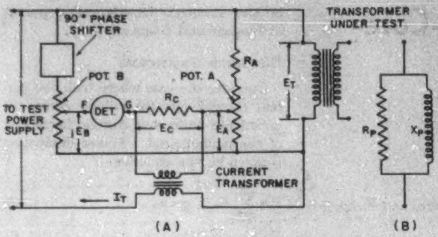


Fig. 1. Basic circuit of the core-loss bridge

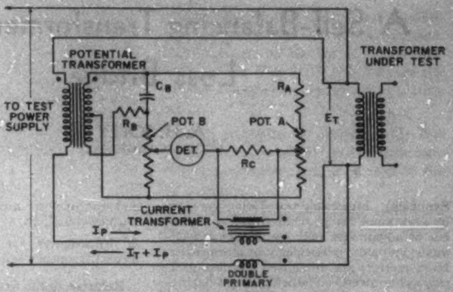


Fig. 2. Example of practical circuit synthesis for the core-loss bridge. The double primary on the current transformer permits tare compensation

Basic Bridge Circuit

Fig. 1(A) shows a form of co-ordinate potentiometer bridge and is the basic circuit chosen for transformer-core-loss measurements. Although a current transformer is shown in the current circuit, the appendix shows that a mutual inductor may also be employed. This circuit measures the equivalent parallel resistance, R_p, and reactance, X_p, as indicated in Fig. 1(B).

Two potentiometer systems are shown connected across the terminals of the transformer under test. Across potentiometer *A* is a fraction of the test voltage (E_t) which is exactly in phase with E_t. Across potentiometer *B* is a fraction of E_t exactly 90 degrees out of phase (leading if a current transformer is employed) with E_t. Resistor R_a and the 90-degree phase shifter in the *A* and *B* potentiometer circuits merely symbolize the conditions required in these branches. The designer is at liberty to synthesize these circuits in any manner he may choose provided only that the basic phase and magnitude requirements are met.

In operation, test current (I_t) flows through the excitation winding of the transformer under test and through the current transformer in series. A voltage (E_c) proportional to I_t and in phase with I_t is developed across the current transformer burden resistance (R_c). E_c is complex and consists of real and quadrature components relative to the applied test voltage E_t. For the present, only voltages and currents at fundamental frequency will be considered. In transformer testing harmonics of the fundamental are present in abundance in the voltage E_c. The matter of harmonics in this portion of the circuit is treated in the discussion of the tuned preamplifier.

When the bridge is balanced, portions of the voltages across potentiometers *A* and *B* are chosen to construct a phasor exactly equal in magnitude, but opposite in phase, to E_c. The net voltage appearing in the detector loop is then zero and is indicated by zero voltage at detector terminals *f* and *g*. The detector loop is considered as the path external to the detector terminals through R_c and those portions of potentiometer *A* and *B* on the lower sides of the sliders as shown in Fig. 1(A).

It is shown in the appendix that at balance the effective parallel resistance of the impedance under test is

$$R_p=\frac{k_c}{k_a k_1} \tag{5}$$

and the effective reactance is

$$X_p=\frac{k_c}{k_b k_2} \tag{6}$$

where k_1 and k_2 are those fractions of potentiometers *A* and *B*, respectively, included in the detector loop while k_a, k_b, and k_c are preset constants as defined in the appendix.

Determination of Core Loss

In determining transformer core loss the operator is not directly concerned with values of R_p and X_p. In fact, for core loss alone the value of X_p is of no concern at all. It is necessary, however, in operation of the bridge, to balance the *B* potentiometer if the precision with which the *A* potentiometer must be set is not to be impaired.

Core loss is determined by applying equation 1 which conveniently is rewritten as

$$\text{Core loss watts}=P=\frac{E_n^2}{R_p} \tag{7}$$

where E_n is the nominal voltage at which core loss is specified. The test voltage E_t will generally be set close to E_n so that (due to the nonlinear nature of the transformer) R_p will be presented to the bridge very near the correct level for E_n. It is shown in the discussion of the effect of voltage variations that E_t does not have to be set with great accuracy when bridge methods are employed. The use of E_n in the loss computation assures maximum accuracy in determining core loss.

Substituting equation 5 in equation 7 yields

$$P=\frac{E_n^2 k_a k_1}{k_c} \tag{8}$$

The core loss is directly proportional to k_1. By suitable choices of the constants k_a and k_c, the dial on potentiometer *A* can be made direct reading for any particular value of E_n. If potentiometer *A* is linear, the watts loss scale will be linear.

Design for Computation of Per-Cent Core Loss and Direct Readings

In transformer-core-loss testing certain conveniences are offered by expressing the results in terms of per-cent core loss relative to the kilovolt-ampere (kva) rating of the transformer under test. This is a form of normalization which permits a direct comparison of the losses in transformers of all kva ratings. A further convenience is offered by permitting the use, generally, of but a single limited scale on the loss dial though the bridge may be designed to measure losses over a wide range of kva ratings.

Another convenience is provided when a measuring instrument reads directly in terms of the desired data thereby eliminating subsequent computations. All of these features can be realized in the core-loss bridge by suitable design.

DEVELOPMENT OF DESIGN PROCEDURE

Equation 8 expresses core loss in watts. Expressing the loss as per cent of kva rating F,

$$\% \text{ core loss} = \frac{100\ P}{1{,}000\ F} = \frac{E_n^2 k_a k_1}{10 k_c F} \tag{9}$$

It is shown in the appendix that, for reasons favoring self-balancing of the bridge, equation 9 is more convenient in the form

$$\% \text{ core loss} = \frac{E_n k_a k_1}{10 k_c \left(\dfrac{F}{E_n}\right)} \tag{10}$$

The form of equation 10 establishes a mode of operation for the bridge which will become clearer as this discussion proceeds.

With the exception of k_1, all factors in both the numerator and denominator of equation 10 may be regarded as constants for any particular transformer. This is a matter of practicality for, by presetting these so-called constants (which are functions of the kva and nominal voltage ratings of the transformer to be tested), the range of transformer losses is, in effect, anticipated.

Dividing both sides of equation 10 by k_1 yields the expression

$$K_r = \frac{\% \text{ core loss}}{k_1} = \frac{E_n k_a}{10 k_c \left(\dfrac{F}{E_n}\right)} \tag{11}$$

K_r may be termed the loss dial constant. It will be recalled that $0 < k_1 < 1$. Therefore, to determine the dial constant for a particular bridge design, it is merely necessary to assume $k_1 = 1$ and the dial constant becomes numerically equal to the maximum value of the desired per-cent loss scale on the bridge. This simply means that the slider on potentiometer A will be set at its maximum position ($k_1 = 1$) when measuring this maximum value of core loss. Equation 11 is an important design equation.

Choosing the full-scale value of per-cent core loss is the first step in designing the bridge. Next, an arbitrary choice of reference values for E_n, F, and k_c must be made. These are designated as E_{no}, F_o, and k_{co}, respectively. The determination of k_{co} is made by first deciding what practical voltage, E_c, is required across the current transformer burden resistance, R_c. Often this will be in the range between 1 and 10 volts. Since

$$E_c = I_t k_c \tag{12}$$

it follows that

$$k_{co} = \frac{E_c}{I_{to}} \tag{13}$$

where I_{to} may be taken as a typical value of exciting current for a transformer having the chosen reference ratings of E_{no} and F_o. Choice of I_{to} is not critical but once a value for k_{co} has been established, k_{co} must not be changed throughout the design.

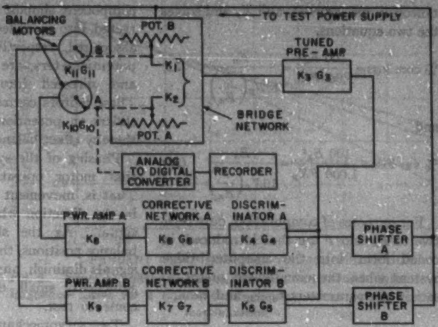


Fig. 3. Block diagram of the dual servomechanism employed for self-balancing

These reference data now permit the establishment of a constant reference value (D_o) for the denominator of equation 11.

$$D_o = \frac{10 k_{co} F_o}{E_{no}} \tag{14}$$

Equation 11 can be rewritten as the following

$$K_r = \frac{E_n k_a}{D_o} \tag{15}$$

from which, since K_r is a constant,

$$k_a = \frac{K_r D_o}{E_n} \tag{16}$$

Note that, as treated here, k_a is a function of only one variable, E_n, K_r and D_o having been established as constants.

The value of D_o, once established for a given bridge design, remains unchanged for any transformer which might be tested. Equating the denominator of equation 11 to D_o,

$$\frac{10 k_c F}{E_n} = D_o \tag{17}$$

from which

$$k_c = \frac{D_o E_n}{10 F} = k_{co} \left(\frac{F_o}{F}\right)\left(\frac{E_n}{E_{no}}\right) \tag{18}$$

The chosen mode of operation makes k_c a function of two variables, E_n and F. The remaining factors in equation 18 have been established as design constants.

This completes the procedure for the basic design of the loss branch of the bridge. Similar treatment for the reactance branch, containing potentiometer B, yields the design equations which follow.

For potentiometer B the reactance dial constant, which is analogous to K_r in equation 11 and is numerically the full-scale value of the per-cent var (reactive volt-ampere) loss (so-called) scale, is

$$K_x = \frac{\% \text{ var loss}}{k_2} = \frac{E_n k_b}{10 k_c \left(\dfrac{F}{E_n}\right)} \tag{19}$$

It is to be noted that the denominator of equation 19 is identical with that of equation 11. Therefore, k_c can have but one value and is given by equation 18. It follows that D_o remains unchanged and that

$$k_b = \frac{K_x D_o}{E_n} \tag{20}$$

which is analogous to equation 16.

Dial constants K_r and K_x need not have identical values. That is, the per-cent core loss and per-cent var loss ranges do not have to be equal. Where the power factor of the transformer under test is less than 70.7%, the var loss will be greater than the core loss. For such cases K_x might be chosen larger than K_r.

Following the procedure outlined and having established numerical values for k_a, k_b, and k_c, the designer may now proceed to synthesize the circuits associated with these parameters. Much freedom is afforded in these areas. An example of synthesized circuit design is shown in Fig. 2. A method of tare compensation is also shown.

REGARDING THE BRIDGE AS A COMPUTER

As treated in this paper the core-loss bridge is a measuring instrument and a form of analog computer which simultaneously measures R_p and X_p and solves the two equations,

$$\% \text{ core loss} = \frac{100\, E_n^2}{1{,}000 F R_p} = \frac{E_n}{10 R_p \left(\frac{F}{E_n}\right)} \tag{21}$$

and

$$\% \text{ var loss} = \frac{100\, E_n^2}{1{,}000\, F X_p} = \frac{E_n}{10 X_p \left(\frac{F}{E_n}\right)} \tag{22}$$

For purposes of computation, data relative to E_n and F are thus introduced, in coded form, into the computer-bridge system when the circuits are preset to satisfy the parameters k_a, k_b, and k_c.

Self-Balancing

As described in the foregoing sections, the core-loss bridge could just as well be manually balanced. However, if the application is in a system of automation, self-balancing is necessary. Also, some form of automatic read-out is essential. A relatively high output torque generally is required for operation of analog to digital converters, limit-sensing devices, or to perform other tasks requiring appreciable energy.

Servomechanism techniques are applicable here. These techniques, particularly for instrument servomechanisms, have been well established, and it is not necessary to discuss the subject in any great detail.[5,6] Only those parts of the system peculiar to self-balancing a-c impedance bridges are treated.

DUAL SERVOMECHANISM

Fig. 3 shows, in block diagram form, the arrangement whereby simultaneous balances for both core loss and var loss have been accomplished. A dual system is employed to make the two balances. These are carrier or all-a-c types of servomechanisms.[5] This is natural since the impedance measurements are made with alternating current and the bridge detector or error signals are a-c. Hence the carrier frequency is that of the test voltage applied to the impedance under test. When the bridge is balanced the carrier is suppressed.

With reference to Fig. 3, the output at the bridge detector points (points f and g, Fig. 1) is the error signal for the dual servomechanism. This signal is fed into a tuned preamplifier and hence into two phase discriminators. The function of the discriminators is to separate the error signal coming through the preamplifier into component parts resulting from unbalanced positions of the sliders on potentiometers A and B. The component signals, having now been directed into so-called A and B channels associated, respectively, with A and B potentiometers, are separately amplified and directed into balancing motors. These motors control the position of the sliders on potentiometers A and B and thereby effect balance of the bridge.

Phasing of the system is always such that motor operation is degenerative. That is, movement of a slider is always in the direction which reduces the error signal. As the sliders approach the balance positions, the corresponding error signals diminish, and when the errors are sufficiently small, the balancing motors come to rest.

Two servomechanisms are thus superimposed on the bridge (the servomechanism summing device) and the preamplifier (the error detector). These two systems operate essentially independently of each other, the degree of independence being determined by the accuracy with which separation takes place in the phase discriminators.

LOOP TRANSFER FUNCTIONS

Although it is beyond the scope of this paper to present the complete design of the servomechanisms employed for self-balancing the core-loss bridge, some discussion of the makeup of the loops and the transfer functions of certain of the loop components is included in the following.

A servomechanism loop is characterized by an over-all transfer function which describes the open loop gain and phase shift, important factors in determining system performance and stability.[5,6] Each component of a loop is characterized by an individual transfer function. It is the continued product of all component transfer functions that makes the total open-loop transfer function.

In Fig. 3, considering the loop associated with potentiometer A as loop A, the individual components and transfer functions are, beginning with the bridge network:

1. Core loss arm of bridge, K_2.
2. Tuned preamplifier, K_3G_3.
3. Channel A discriminator, K_4G_4.
4. Corrective network, K_5G_5.
5. Power amplifier, K_8.
6. Balancing motor, $K_{10}G_{10}$.

In these transfer functions the K's represent frequency-invariant ratios while the G's, if present, represent complex frequency dependent ratios. These are all gain expressions. That is, each transfer function is the complex ratio of output to input of the component. The transfer function of loop A becomes

$$K_AG_A = K_2K_3K_4K_5K_8K_{10}G_3G_4G_5G_{10} \tag{23}$$

Similarly, the transfer function of loop B becomes

$$K_BG_B = K_1K_3K_6K_7K_9K_{11}G_3G_6G_7G_{11} \tag{24}$$

BRIDGE TRANSFER FUNCTIONS

Bridge transfer functions, K_1 and K_2, need some clarification. If potentiometers A and B are rotary potentiometers, as usually is the case, the inputs to the bridge are in the form of shaft rotations (from the balancing motors). The output of the bridge (strictly as a loop component) is a voltage which appears at the detector terminals. The bridge transfer functions, one for each of the two potentiometer sections, are ratios of incremental voltages out (ΔE_o) to incremental input shaft angles ($\Delta\theta_i$). It is shown in the appendix that, for the mode of operation established under "Design Procedure," K_1 and K_2 are constants and may be computed from the relations

$$K_1 = \frac{\Delta E_{ob}}{\Delta\theta_{tb}} = \frac{E_{pb}}{\theta_{tb}} \tag{25}$$

and

$$K_2 = \frac{\Delta E_{oa}}{\Delta\theta_{ta}} = \frac{E_{pa}}{\theta_{ta}} \tag{26}$$

where E_{pa} and E_{pb} are total voltages across potentiometers A and B and θ_{ta} and θ_{tb} are the total effective rotation angles for these potentiometers. Values computed from equations 25 and 26 may be employed directly in equations 23 and 24.

TUNED PREAMPLIFIER

The presence of relatively large amounts of harmonic voltages in the error signals from the bridge makes some form of discriminative filtering absolutely necessary in the core-loss bridge. There are many circuit arrangements for accomplishing this, including the use of selective feedback in the preamplifier.[7] In transformer-core-loss testing it frequently is found necessary to have as much as 70 or 80 decibels of discrimination against third and higher harmonics. That is, the preamplifier gain for third and higher harmonics may have to be less than $^{1}/_{1{,}000}$th of that for the fundamental frequency.

Filtering with this degree of selectivity often results in so-called side band attenuation. For the case at hand the carrier frequency is that of the fundamental frequency at which tests are made, such as 60 cps. In the system under

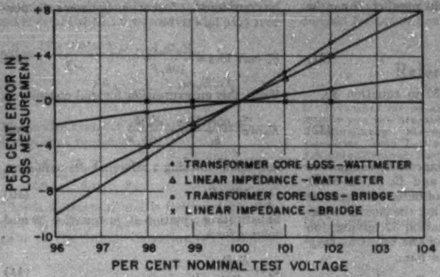


Fig. 4. Comparison of voltage errors for wattmeter and bridge methods of measurements on a linear impedance and a typical transformer

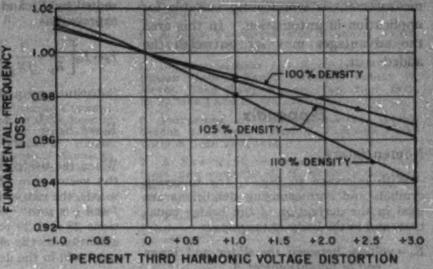


Fig. 5. Effect of voltage waveform variations on total and fundamental frequency losses for a typical transformer at various levels of core magnetization

consideration here, side-band frequencies extending from 20 cps below to 20 cps above the fundamental may have to be considered.

It may be assumed, therefore, that the range of carrier modulation frequencies effectively extends from zero to approximately 20 cps. It is over this frequency range that the characteristics of the preamplifier must be considered. Typically, a tuned amplifier may have a transfer function approximated by the equation,

$$K_3G_3 = \frac{K_3}{1+j\omega_m T_3} \tag{27}$$

where K_3 is the zero frequency (unmodulated carrier) gain, ω_m is the angular frequency (radians per second) of the modulating signal, and T_3 is the effective time constant (seconds) of the amplifier or filter system.

The significance of equation 27 is that a lagging phase shift occurs in the preamplifier and approaches 90 degrees when $\omega_m T_3 >> 1$. At the same time the absolute gain diminishes with increasing ω_m. Equation 27 may be used directly in equations 23 and 24 if it characterizes the tuned preamplifier chosen for the system.

Typical transfer functions for the remaining components of the loops can be found in the references.[4,6,8] When considering the balancing motors it should be remembered that these are employed as positioning devices and the corresponding transfer functions must be employed in equations 23 and 24.

Bridge Characteristics and Performance

Effect of Test Voltage Variations

In the development of the impedance bridge equations in the appendix it is shown that, for linear impedances, the test voltage cancels out of the balance equations. Therefore, in determining power losses in linear impedances by bridge methods, variations in test voltages do not directly introduce errors. This is illustrated in Fig. 4.

For comparison, the effect of voltage variations on loss measurements by wattmeter methods is shown. Here a 1% error in setting the test voltage results in an error of approximately 2% in the loss measurement.

Fig. 4 also shows a comparison of the effects of test-voltage variations when core loss in a typical transformer is measured by the two methods. Though the bridge method shows some dependency upon the value of the test voltage held, the error is considerably less than when a wattmeter is employed. For a 1% error in setting the test voltage, the bridge shows an error of about 0.5%, whereas the wattmeter shows an error of about 2.6%. For core-loss measurement by bridge methods, test voltage settings are approximately one fifth as critical as when employing a wattmeter.

Effect of Voltage Waveform Variations

Early in this paper it was pointed out that impedance bridges yield results in terms of fundamental frequencies only. Fig. 5 shows how, for a typical transformer, the ratio of total loss to fundamental frequency loss is affected by variations in third harmonic voltage distortion. The data for the three curves shown in Fig. 5 were determined by comparing simultaneous readings taken with a precision wattmeter and a carefully calibrated impedance bridge. As was expected it was quickly determined that third harmonic voltage distortion accounts for the major portion of the differences between total and fundamental frequency losses. It is apparent that the ratio of total to fundamental loss varies essentially linearly with distortion.

Summary

It should not be construed from the arguments set forth in this paper that impedance bridges are in direct competition with wattmeters as loss-measuring instruments. Each class of instruments has its special field of application. Impedance bridges, in general, are considerably more costly than wattmeters and application should be limited to those cases where the advantages of the bridge method offset the added cost.

A partial list of advantages offered by the impedance bridge, with or without the self-balancing feature, includes the following

1. High accuracy.
2. High sensitivity.
3. Direct reading.
4. Test-voltage setting is less critical.
5. Can cover a very wide range in a single instrument.
6. Can perform certain analog computations for data reduction.

When the self-balancing feature is included these additional advantages are provided:

1. Fast operation; balance in 2 seconds with 0.1% precision can be accomplished.
2. Can provide very high mechanical output torque for performing auxiliary tasks such as recording, test-limit sensing, or sorting.
3. Eliminates human operating and reading errors.
4. Can be completely and remotely controlled with punched cards, punched tape, or other automatic programming equipment.

The impedance bridge method of loss

measurement is emminently suitable for application in automation. In this area the advantages may far outweigh the added cost.

Appendix

Nomenclature

With reference to Fig. 1, the following symbols and corresponding definitions are used in the derivation of the bridge equations and in the body of the paper:

E_a = voltage between the slider and lower end of potentiometer A
E_b = voltage between the slider and lower end of potentiometer B
E_c = voltage across the current transformer burden resistance (R_c)
E_n = nominal voltage at which loss is specified
E_t = test voltage applied to the impedance under test
E_{pa} = total voltage across potentiometer A
E_{pb} = total voltage across potentiometer B
F = kva rating of transformer under test
F_o = kva rating of a transformer chosen for design reference
I_t = test current due to E_t
$k_a = E_{pa}/E_t$, the fraction of E_t appearing across potentiometer A
$k_b = E_{pb}/E_t$, the fraction of E_t appearing across potentiometer B
$k_c = E_c/I_t$, the factor relating I_t to E_c
k_{co} = a design reference value of E_c/I_t
k_1 = the fraction of potentiometer A included between the slider and lower end
k_2 = the fraction of potentiometer B included between the slider and lower end
K_r = per-cent core loss/k_1 (a dial or scale constant for potentiometer A)
K_2 = per-cent var loss/k_2 (a dial or scale constant for potentiometer B)
K_1 = transfer function for potentiometer B
K_2 = transfer function for potentiometer A
R_p = effective parallel resistance of impedance under test
X_p = effective parallel reactance of impedance under test
ΔE_o = incremental output voltage presented to the detector terminals when the bridge is unbalanced
Δk_1 = an increment added to k_1
θ_{ta} = effective total rotational angle for potentiometer A

Derivation of Bridge Balance Equations

Referring to Fig. 1(A), the voltage appearing between the slider and lower end of potentiometer A is

$$E_a = k_a k_1 E_t \qquad (28)$$

The voltage between the slider and lower end of potentiometer B, being 90 degrees out of phase (leading) with respect to E_t, is the equation

$$jE_b = jk_b k_2 E_t \qquad (29)$$

The voltage appearing across R_c is

$$E_c = k_c I_t \qquad (30)$$

Since the equivalent circuit of the transformer or impedance under test is represented by R_p and X_p in parallel, I_t can be expressed as

$$I_t = E_t\left[\frac{1}{R_p} + \frac{1}{jX_p}\right] \qquad (31)$$

Substituting equation 31 in equation 30,

$$E_c = k_c E_t\left[\frac{1}{R_p} + \frac{1}{jX_p}\right] \qquad (32)$$

When the bridge is balanced E_c is equal to the vector sum of E_a and E_b or, in other words, the net voltage of detector terminals f and g is zero. Equating the sum of equations 28 and 29 to equation 32, giving due attention to the signs of the terms as they must add in the detector loop,

$$k_a k_1 - jk_b k_2 = k_c\left[\frac{1}{R_p} + \frac{1}{jX_p}\right] \qquad (33)$$

Note that, at this point, the voltage, E_t, has cancelled out of equation 33, the principal balance equation. Separating the real and imaginary terms of equation 33 and solving for R_p and X_p,

$$R_p = \frac{k_c}{k_a k_1} \qquad (34)$$

and

$$X_p = \frac{k_c}{k_b k_2} \qquad (35)$$

Use of a Toroid in Place of a Current Transformer

Equations 34 and 35 are based on the use of a current transformer as shown in Figure 1(A). If a toroidal mutual inductor is employed in place of the current transformer and the resistance R_c is omitted, E_c is advanced in phase by 90 degrees. Setting up the bridge equations to take this into account and solving for R_p and X_p,

$$R_p = \frac{k_c}{k_b k_2} \qquad (36)$$

and

$$X_p = \frac{k_c}{k_a k_1} \qquad (37)$$

It is seen that where a current transformer is employed, R_p is determined from the setting of potentiometer A. When a toroidal mutual inductor is substituted, R_p then is determined from potentiometer B. The same basic network is therefore suitable for use with either a current transformer or a toroidal mutual inductor, due attention being given to polarities in the detector loop.

The detector loop is considered as the complete path external to the detector through R_c (or the mutual inductor, as the case may be) and through those portions of the two potentiometers which are below their sliders.

Bridge Sensitivity

In determining the bridge sensitivity to incremental adjustments of the sliders on the potentiometers, it is first necessary to determine the voltage across the potentiometers for the chosen mode of operation of the bridge. To simplify the analysis only the sensitivity relative to potentiometer A will be considered. Equation 10 shows the percent core loss to be

$$\%\text{ core loss} = \frac{E_n^2 k_a k_1}{10 k_c F} \qquad (38)$$

Using this an expression for dial constant is

$$K_r = \frac{\%\text{ core loss}}{k_1} = \frac{E_n^2 k_a}{10 k_c F} \qquad (39)$$

and upon choosing a reference denominator,

$$D_o = 10 k_{co} F_o \qquad (40)$$

Substituting equation 40 in equation 39 and solving for k_a,

$$k_a = \frac{K_r D_o}{E_n^2} \qquad (41)$$

The voltage across potentiometer A becomes

$$E_{pa} = k_a E_t = k_a E_n = \frac{K_r D_o}{E_n} \qquad (42)$$

E_a, the voltage between the slider and lower end of potentiometer A, is

$$E_a = k_1 E_{pa} = \frac{k_1 K_r D_o}{E_n} \qquad (43)$$

Suppose an increment (Δk_1) is now added to k_1. E_a then becomes

$$E_a' = \frac{(k_1 + \Delta k_1) K_r D_o}{E_n} \qquad (44)$$

If the bridge was balanced for equation 43, the output (ΔE_o) at the detector terminals, upon adding the increment Δk_1, is now

$$\Delta E_o = E_a' - E_a = \frac{\Delta k_1 K_r D_o}{E_n} \qquad (45)$$

With this approach ΔE_o is a function of E_n and therefore not a constant for the operational mode established by equation 38. It is highly desirable for the bridge sensitivity to be constant. Equation 11 is a rearrangement of equation 10 and shows

$$\%\text{ core loss} = \frac{E_n k_a k_1}{10 k_c\left(\frac{F}{E_n}\right)} \qquad (46)$$

Following the procedure outlined but using equation 46 instead of equation 38 yields

$$\Delta E_o = \Delta k_1 K_r D_o \qquad (47)$$

This provides the desired constant relationship between Δk_1 and ΔE_o and shows why the form of equation 46 is preferable to that of equation 38.

Bridge Transfer Functions

Let potentiometer A have an effective total rotational angle (θ_{ta}). Then for any value of k_1 the slider is at an angle

$$\theta_{1a} = k_1 \theta_{ta} \qquad (48)$$

from which $k_1 = \dfrac{\theta_{1a}}{\theta_{ta}}$ (49)

Adding an increment of angle ($\Delta\theta_{1a}$) to θ_{1a}, it follows that

$$k_1 + \Delta k_1 = \frac{\theta_{1a} + \Delta\theta_{1a}}{\theta_{ta}} \qquad (50)$$

The incremental output voltage at the detector terminals then becomes, following the mode established by equation 48,

$$\Delta E_0=\left(\frac{\theta_{1a}+\Delta\theta_{1a}}{\theta_{ta}}\right)K_rD_0-\left(\frac{\theta_{1a}}{\theta_{ta}}\right)K_rD_0=\frac{\Delta\theta_{1a}K_rD_0}{\theta_{ta}} \qquad (51)$$

The bridge transfer function for potentiometer *A* then becomes

$$K_2=\frac{\Delta E_0}{\Delta\theta_{1a}}=\frac{K_\theta D_0}{\theta_{ta}}=\frac{E_{pa}}{\theta_{ta}} \qquad (52)$$

This is a constant and is independent of k_1 and E_n. It provides for a constant loop gain in the bridge-balancing servomechanism. A similar expression for K_1 can be derived for potentiometer *B*.

References

1. INVESTIGATION OF AN ALTERNATING CURRENT BRIDGE FOR THE MEASUREMENT OF CORE LOSS IN FERROMAGNETIC MATERIALS AT HIGH FLUX DENSITIES, Irvin L. Cooter, William P. Harris. *Journal of Research*, National Bureau of Standards, Washington, D. C., vol. 57, no. 2, Aug. 1956, pp. 103–12.

2. ALTERNATING CURRENT BRIDGE METHODS (book), B. Hague. Sir Isaac Pitman and Sons, Ltd., London, England, 1943.

3. A SELF-BALANCING CAPACITANCE BRIDGE, A. H. Foley. *AIEE Transactions*, vol. 66, 1947, pp. 797–801.

4. A FLUX VOLTMETER FOR MAGNETIC TESTS, G. Camilli. *Ibid.*, vol. 45, 1926, pp. 721–28.

5. SERVOMECHANISMS AND REGULATING SYSTEM DESIGN (book), Harold Chestnut, Robert W. Mayer. John Wiley & Sons, Inc., New York, N. Y., vols. I and II, 1951, 1955.

6. PRINCIPLES OF SERVOMECHANISMS (book), Gordon S. Brown, Donald P. Campbell. John Wiley & Sons, Inc., 1948.

7. A NEW TYPE OF SELECTIVE CIRCUIT AND SOME APPLICATIONS, H. H. Scott. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 26, Feb. 1938, pp. 226–35.

8. PHASE SELECTIVE DETECTORS, Curtis R. Schafer. *Electronics*, New York, N. Y., Feb. 1954, pp. 188–90.

Development of a Thermal-Neutron Flux-Measuring Instrument

C. V. WEAVER
NONMEMBER AIEE

C. K. SMITH
NONMEMBER AIEE

J. W. CHASTAIN
NONMEMBER AIEE

Synopsis: The instrument measures thermal-neutron flux directly and automatically. Measurements are nearly independent of gamma flux and ambient temperature. The method depends upon balancing the temperatures of an electrically heated balancing element and a fission-heated uranium-fueled sensing element. The flux-detector assembly measures only 3/16-inch OD (outside diameter) by $2^1/_4$ inches long.

THE OBJECTIVE of this research program was to develop an instrument to measure thermal-neutron flux directly and accurately in the range from 10^{13} to 10^{14} neutrons/cm^2 (square centimeters)/second. The desired detector life was to be about 100 hours at an average flux of 5×10^{13} neutron/cm^2/second at an ambient temperature of approximately 600 F (degrees Fahrenheit). In addition, the detector assembly was to be movable over a 20-foot length inside a 0.210-inch ID (inside diameter) tube. The detector was to be operable over the bottom 4 feet of the total travel.

To avoid the use of correction factors and curves and therefore make the instrument convenient to use, it was initially decided that the instrument should be self-compensating for ambient temperature changes and should not include gamma flux in its measurements.

Analysis of Measuring Methods

Two basic methods of neutron-flux measurements were considered: (1) the usual ion-chamber techniques, and (2) the use of the fission heat generated by neutrons in a uranium-fueled element. A variation of the latter method was finally chosen; the reasons for this choice are given in the following discussion.

ION-CHAMBER METHOD

The smallest ion chamber described in the literature was 1/2-inch OD and contained a center electrode which was only slightly smaller than the total allowable outside diameter of the instrument to be developed in this program. It appeared that the problem of reducing the size of an ion chamber to less than 0.210-inch OD while maintaining the necessary potentials would be extremely difficult of solution.

In addition, the 600 F ambient temperature requirement imposed still more formidable materials problems. In view of the greater promise shown by the fission-heat method, the ion-chamber method was abandoned.

If two identical elements, one containing enriched and one containing natural uranium are placed together in a neutron flux, a temperature difference will exist between the two. Both will be heated equally by the gamma radiation, so that the difference in the amount of heat produced in the two elements is the extra heat produced by fission in the element containing enriched uranium.

The temperature rise caused by fission in the fueled element above the reference temperature of the unfueled element can be measured by the potential generated in a differential thermocouple between them. This potential could be used as a measure of the neutron flux if it were not for the fact that the heat loss is not a linear function of the heat generated. Hence, the temperature differential measured by the thermocouple will be a function of the temperature, and the system will not be self-compensating for temperature changes.

However, if the element containing natural uranium is heated electrically to the temperature of the fueled element, both will lose heat at the same rate. The electric power required to accomplish this is equal to the difference in heat generation between the elements, and is proportional to the neutron flux. In any ambient temperature range where the product of the macroscopic fission cross section and the neutron velocity can be considered constant, the electric power supplied is directly proportional to the neutron density. Because thermal flux is defined as neutron density multiplied by a mean neutron velocity of 2,200 meters per second, the electric power supplied is also directly proportional to this thermal flux.

Since this electric power could be calibrated directly as a linear function of thermal neutron flux with very small error, and since the possibility existed that this power could be matched accurately and automatically to the fission power by a reasonably simple control, this system was chosen for development.

Paper 57-406, recommended by the AIEE Nucleonic and Radiation Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE 2nd Nuclear Engineering and Science Conference, Philadelphia, Pa., March 11–14, 1957. Manuscript submitted April 17, 1957; made available for printing September 10, 1957.

C. V. WEAVER, C. K. SMITH, and J. W. CHASTAIN are with Battelle Memorial Institute, Columbus, Ohio.

This work was performed under Atomic Energy Commission Contract no. *W-7405-eng-92*.

Table I. Results of Design Calculations*

ϕ, Neutron/ Cm^2/Second	P_S, Watts	P_B, Watts	P_S-P_B, Watts	T, F
5×10^{14}	36.200	0.276	35.900	2,550
1×10^{14}	7.230	0.055	7.180	1,650
5×10^{13}	3.620	0.028	3.590	1,400
1×10^{13}	0.723	0.005	0.718	850
5×10^{12}	0.362	0.002	0.359	750
1×10^{12}	0.072	0.001	0.072	

* See Appendix I for sample calculations.

ϕ = thermal neutron flux

P_S = power generation by fission in sensing element

P_B = power generated by fission in balancing element

P_S-P_B = electric power required to balance fission power $(P_S-P_B)=7.18\times10x^{14}\ \phi$ watts

T = calculated temperatures of elements assuming emissivities of only 70% for elements and capsules, and an ambient or well temperature of 600°F

Design Considerations

The accuracy of the power measuring method of determining fission rates depends upon the elements losing heat at the same rate. Since the thermal conductivity of materials can be affected by radiation damage and since either forced or natural convection is impractical or uncertain, radiation seemed the best method of removing the heat generated in both elements.

Four main factors determined the size and configuration of the flux measuring elements:

1. To insure heat loss by radiation, a reasonable gap must exist between the elements and the 3/16-inch OD containing case.
2. The thermocouple beads should be located at the midpoint of each element to insure maximum measuring accuracy.
3. In the small space allowed, it would be impractical to place the elements side by side. Furthermore, with such an arrangement, the elements would partially shield one another from the flux.
4. The active portion of the device should be made as compact as possible to keep both elements in essentially the same flux.

These requirements were satisfied by making the elements hollow cylinders 3/32-inch OD, 1/32-inch ID, 1/2 inch long and arranging them in tandem.

One of the most important considerations in choosing the element material was the electrical resistivity. To keep the heating current small so that a negligible amount of heating takes place in the power leads, a resistance of 100–150 ohms would keep the heating potential below 50 volts which would simplify insulation problems.

Initially several high-temperature alloys were investigated for use in the elements. Wires of uranium-thorium and uranium-zirconium were of interest because of their radiation damage stability and their fairly high resistivity. However, to obtain resistance of 100 ohms, several inches of 0.002-inch diameter wire is required. This wire could be wound on a mandrel 1/2 inch long and 3/32 inch in diameter. However, because of the difficulties of drawing and winding the wire, these materials were abandoned in favor of more promising ones.

Many ceramic materials can be extruded in the proper size and shape to serve as elements. The resitivity of some of these materials can be adjusted to 3 to 5 ohms-centimeters which will give resistances of 100 to 150 ohms for the element. In addition, materials with specific resistivities of this order have very small temperature coefficients. The disadvantages of these materials are their unknown radiation-damage susceptibility and in many cases poor mechanical strength.

The feeling that these disadvantages could be overcome by a suitable choice of materials and fabrication techniques prompted a first-line effort. A material with the proper resistivity and good mechanical strength containing 49.18 w/o (per cent by weight) Al_2O_3, 49.18 w/o $MoSi_2$, and 1.6 w/o UO_2 was developed. Later tests showed that radiation damage was negligible within the required flux and time conditions.

Calculations were made to determine the amount of uranium-235 required in the sensing element to produce about 7 watts of power. This power level was decided upon primarily as a compromise between element-temperature considerations and the desire to minimize error due to heat losses in the leads. This power which could be supplied easily electrically to the balancing element was chosen to correspond to a flux of 1×10^{14} neutrons/cm^2/second. Sample calculations are given in Appendix I. The results of these calculations are shown in Table I. The 1.64 w/o uranium-235 will produce 7.18 watts of power in the sensing element at 1×10^{14} neutrons/cm^2/second. Assuming emissivities of only 70% for the surface of the elements and the container case, and an ambient or well temperature of 600 F, the element temperature was calculated to be 1,650 F at this power level.

The present instrument is designed to operate over the flux range between 1×10^{12} and 1×10^{14} neutrons/cm^2/second. The upper limit of operation is determined by the melting point of the silver contacts on the end of the balancing element and by the electrical stability of the ceramic elements at high temperatures. The safe element temperature is estimated to be around 1,650 F which corresponds to a maximum flux of 1×10^{14} neutrons/cm^2/second. The lower limit of operation is determined by the relatively small temperature differences between the two elements at low power level. The instrument can be made to perform equally well in either higher or lower flux ranges by adjusting the amount of enriched UO_2 in the sensing element accordingly.

The estimated maximum instrument error is between 3 and 5% over the flux range from 1×10^{13} to 1×10^{14} neutrons/cm^2/second. At the high flux level where the temperature change is larger per unit change in flux, the instrument is more accurate. On the other hand, at flux levels of 1×10^{12} neutrons/cm^2/second or lower, the over-all error may be greater than 5%.

Calculations have shown that following a step change in power from 1 to 7 watts, the element will reach 90% of its ultimate temperature in less than 13 seconds. This corresponds to a change in flux from 1.3×10^{13} to 9.6×10^{13} neutron/cm^2/second. It is estimated that the fluxmeter should be able to reach an equilibrium within the accuracy of the instrument in 20 seconds or less, after going from a position of minimum to maximum flux, providing that the response time of the power balancing unit matches that of the element assembly. Cooling time about equals heating time.

The useful life of the sensing element assembly is estimated to be about 100 hours at an average flux of 5×10^{13} neutrons/cm^2/second. By this time, approximately 1 w/o of the U-235 will be burned up. Further use of the sensing elements beyond this point will decrease the accuracy of the instrument.

Development of Prototype Instrument

Basic Systems Design

A schematic diagram of the improved basic thermal-neutron-flux-measuring system is shown in Fig. 1. It consists of two main units: (1) the flux-sensing-element assembly, and (2) the power-balancing and measuring assembly.

To analyze this system, suppose the sensing-element assembly is moved from a position of low- to one of high-neutron flux. Because the heat transfer and gamma heating characteristics of the ceramic-sensing and balancing elements have been made almost identical, a temperature differential will exist between the two because of the power being generated by fission in the sensing ele-

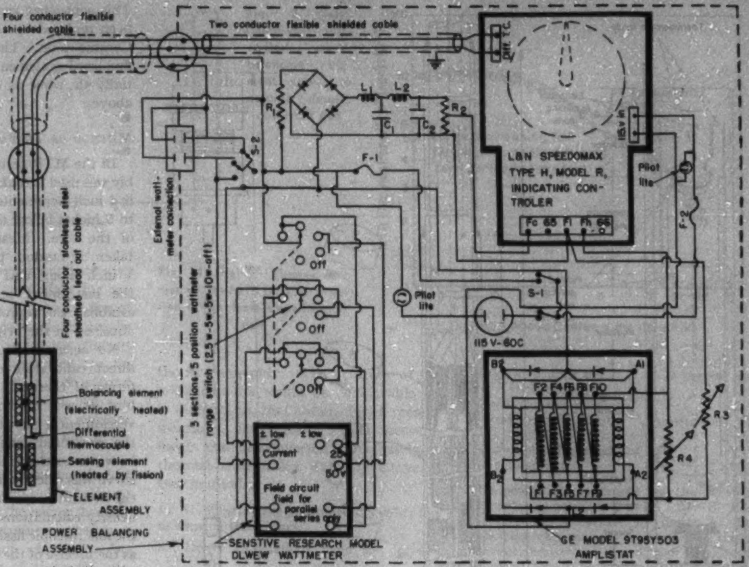


Fig. 1. Basic thermal-neutron-flux-measuring system and circuit diagram

Four conductor flexible shielded cable
Two conductor flexible shielded cable
Four conductor stainless-steel sheathed lead out cable
External wattmeter connection
3 sections-5 position wattmeter range switch (2.5w-5w-5w-10w-off)
R1
L1
L2
C1
C2
R2
S-2
F-1
Off
Off
Off
Pilot lite
S-1
115 V-60C
Balancing element (electrically heated)
Differential thermocouple
Sensing element (heated by fission)
ELEMENT ASSEMBLY
POWER BALANCING ASSEMBLY
± low
± low
Current
25
50v
Field circuit field for parallel series only
SENSTIVE RESEARCH MODEL DLWEW WATTMETER
Diff. T.C.
115 v in
L&N SPEEDOMAX TYPE H, MODEL R, INDICATING CONTROLER
Pilot lite
F-2
Fc 65 Fl Fh 66
B2
A1
F2 F4 F6 F8 F10
F1 F3 F5 F7 F9
B2
A2
L2
R3
R4
GE MODEL 9T95Y503 AMPLISTAT

ment. The electric potential created in the differential thermocouple actuates a proportioning controller which regulates the output of a magnetic amplifier supplying electric heating power to the balancing element. This automatic adjustment of power in the balancing element continues until the temperature difference between the sensing and balancing elements, and the differential thermocouple potential is zero. When this null point is reached, the equivalent of the net fission power is read directly from a wattmeter.

Design of Sensing-Element Assembly

The thermal-neutron flux-sensing-element assembly is shown in Fig. 2. The elements have a composition of 49.18 w/o Al_2O_3, 49.18 w/o $MoSi_2$, and 1.64 w/o UO_2. Natural UO_2 is used in the balancing elements and highly enriched UO_2 in the sensing elements. The elements are extruded and machined ceramic slugs which have been hydrogen-fired at 2,500 F to give them mechanical strength and electrical stability at operating temperatures. The finished elements are 0.093 inch OD, 0.035 inch ID, and 0.500 inch in length. They have an electrical resistance between 100 and 150 ohms. Electric contacts are baked on the balancing element after painting each end with a high-temperature colloidal silver paint.

The elements are assembled in a stainless-steel case 3/16 inch OD by $2^3/_{16}$ inches long. The inside and outside surfaces of the case are acid-etched and blackened by dipping in molten $K_2Cr_2O_7$ to allow maximum radiant-heat transmission through the walls from the elements inside.

The element assembly is connected to a 1/8-inch OD stainless-steel sheathed 4-conductor Ceramo lead-out cable. This cable, approximatley 30 feet, long, contains two no. 24 Chromel wires which are joined to the Chromel differential thermocouple leads, and two no. 24 copper wires which are jointed to the platinum power leads. The cable terminates in a potted Amphenol cable plug. The Ceramo cable can be coiled in a 4-foot-diameter loop without damage to the conductor insulation.

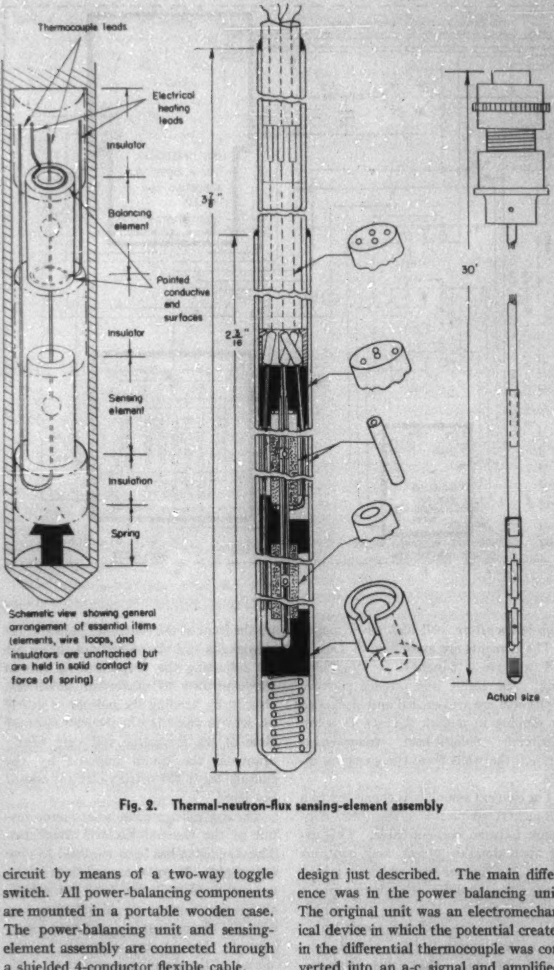
A stainless-steel sleeve joins the case to the cable. The entire unit is evacuated and closed by brazing.

Power Balancing Assembly Design

The principal components and the electrical circuitry of the power balancing unit are shown in Fig. 1. The Leeds and Northrup Speedomax *Type-H* indicating controller is actuated by the signal from the differential thermocouple. The dial on the front of the controller indicates the magnitude and the polarity of the signal. By adjusting the proportioning unit of the controller to maximum sensitivity, that is, by keeping the pointer as nearly on zero as possible, the over-all response time of the fluxmeter will very nearly approach the limits imposed by the thermal lag in the sensing-element assembly itself.

The controller regulates the power output of the General Electric Amplistat. The Amplistat has been modified to give a-c output. It was reasoned that, in the narrow confines of the lead out cables, alternating current would interfere less with the thermocouple signal than direct current. The electric network on the output of the Amplistat provides a feedback which both stabilizes the output against normal line voltage fluctuations and decreases the response time of the Amplistat.

The power supplied to the balancing element is measured by a Sensitive Research special low-power-factor wattmeter. An external switch allows three ranges, 2.5, 5, and 10 watts. The accuracy of the meter is better than 0.5% full scale. Provisions are made for selectively placing either the Research wattmeter or an external recording wattmeter in the power



Schematic view showing general arrangement of essential items (elements, wire loops, and insulators are unattached but are held in solid contact by force of spring)

Fig. 2. Thermal-neutron-flux sensing-element assembly

circuit by means of a two-way toggle switch. All power-balancing components are mounted in a portable wooden case. The power-balancing unit and sensing-element assembly are connected through a shielded 4-conductor flexible cable.

Experimental Tests

In-pile tests were run inside the beryllium reflector of the MTR (Materials Testing Reactor) to evaluate the power balancing method of measuring thermal neutron flux.

EQUIPMENT

The instrument used in the MTR tests was the first model built and differed in certain respects from the more improved design just described. The main difference was in the power balancing unit. The original unit was an electromechanical device in which the potential created in the differential thermocouple was converted into an a-c signal and amplified. The output of the amplifier was supplied to a 2-phase reversible motor which adjusted a variable power transformer in the direction dictated by the polarity of the thermocouple voltage until a null point was reached. While this system worked fairly well, it possessed some inherant disadvantages which led to the use of the present all-electric system.

The long thermocouple and power leads in the original element assembly were threaded through 4-hole ceramic insulators and enclosed in a stainless-steel tube. This somewhat rigid assembly has since been replaced by the much more flexible Ceramo cable. The design of the original flux-sensing-element assembly was essentially the same as the one just described above.

METHOD OF TEST AND CALIBRATION

In the MTR tests, the element assembly was used to make a measuring traverse in 1-inch increments from 16 inches above to 5 inches below the vertical center line of the core. Measurements were then taken at random points in multiples of 1 inch along this length to determine the instrument response time and to establish the reproducibility of the data obtained in the original traverses.

No accurate data were available for direct calibration of the instrument in terms of thermal neutron flux. Therefore, it was necessary to use the conversion curve shown in Fig. 3 to convert the measured balancing-element power in watts to thermal neutron flux. This curve was calculated from the power and the element parameters. In the neutron-density calculations, it was assumed that the macroscopic fission cross section varies as the inverse of the neutron velocity.

TEST RESULTS

Two tests were conducted with identical element assemblies and the same controller used for both. Each element assembly was continuously evacuated while being used.

Analysis of Experimental Data. The experimental curve shown in Fig. 4 was drawn by using the calculated curve from Fig. 3 to convert the average balancing-element power at each position to thermal flux. The latest available MTR flux map of the test location, obtained using foil techniques, was for a core of somewhat different geometry. However, this old curve is shown in Fig. 4 so that basic comparisons can be made.

Using the calculated calibration of power in terms of flux, the maximum deviation of a single instrument reading from the average of all readings at a point where the flux was about 1.04×10^{14} neutrons/cm^2/second was $2^1/_2\%$, and the corresponding deviation for a flux of about 2.2×10^{13} neutrons/cm^2/second was $5^3/_4\%$. If the actual flux at any given point is assumed constant, these represent the maximum instrument errors.

Element Assembly Life. During the first test, the ambient temperature was 850 F instead of the desired maximum 600 F, due to gamma heating of the surrounding equipment. However, the instrument

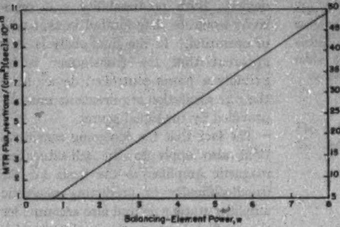


Fig. 3 (left). Calculated conversion curve-balancing element power to neutron density and MTR flux

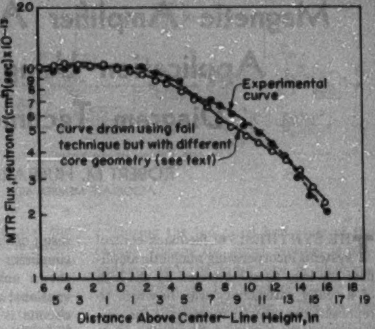


Fig. 4 (right). Thermal-flux map of one MTR reflector position drawn from experimental data

operated at this extreme temperature with excellent reproducibility of data for approximately 6 hours in an average flux of 5×10^{13} neutrons/cm²/second before the electric heating circuit inside the element assembly opened and prevented the recording of additional data. A second assembly was used for about 70 hours in the same average flux with an ambient temperature of approximately 600 F before its circuit shorted.

The assemblies were subsequently lost and could not be examined. It is believed that deterioration of the silver heating contacts due to the extreme temperature caused the open circuit. It is also believed that crumbling of the solid porcelain insulators in the very long lead out tube due to vibration induced by coolant water flow caused the short circuit.

Instrument Response Time. Response time of the instrument was about $1^1/_2$ minutes in going directly from the minimum to the maximum flux region. Measurements in increments up to 3 inches in the test position could be made in less than 1 minute. This shows that a complete core traverse in 1-inch increments can be made in about 1 hour.

Conclusions

The original flux measuring device was capable of measuring the flux distribution at 37 points along a core 3-foot high in a total time of 37 minutes. This is much less time than that required for comparable measurements using other techniques. The maximum instrument error as calculated from the maximum deviation of a single reading from the average of all readings at a given point is between 3 and 5%.

The instruments of improved design which are now being field tested should be capable of making a flux-measuring traverse through a fuel assembly 3-foot long in less than 10 minutes if the sensing unit is moved continuously and slowly through the fuel assembly. The accuracy of the measurements would be comparable to, or better than, that obtained with present measuring techniques. Power measurements can be converted immediately to thermal-neutron-flux data by either of two methods:

1. Calibrating the instrument by other flux measuring techniques.
2. Direct calculations from the formula

$\phi=KP$ where
ϕ=the thermal-neutron flux, neutrons/cm²/second
$P=P_S-P_B$=the power supplied to the balancing element
K=constant (see Table I and calculations of Appendix I)

The power balancing and measuring system are contained in one reasonably portable unit which can be used in any convenient location where 110-volt 60-phase current is available.

Appendix I. Design Calculations for the Thermal-Neutron-Flux-Measuring Instrument

Fission Power Generation

The fission power in the sensing element is:

$$P_S=\Sigma_f\phi V_S$$

where

Σ_f=macroscopic fission cross section of uranium-235
ϕ=thermal neutron flux
V=volume of uranium-235
E=energy released per fission

In terms of the known mass of UO₂ present in each element:

$$P_S=\frac{M_{UO_2}N_O\sigma_f\phi E}{A_{UO_2}}$$

The following apply to this equation.

M_{UO_2}=measured mass of UO₂ present in each element
N_O=Avogadros number
σf=microscopic fission cross-section of U-235
A_{UO_2}=molecular weight of UO₂
E=enrichment

The power P_S generated by the sensing element was calculated to be $7.23\times10^{-14}\phi$.

The power generated by fission in the balancing element, P_B is simply P_S times the enrichment ratio and was calculated to be $5.53\times10^{-16}\phi$ watts. The difference P_S-P_B or $7.18\times10^{-14}\phi$ watts is the amount of power which must be supplied electrically.

Surface Temperature of Elements

Assuming the heat has to be entirely radiant, the surface temperature of the sensing element was calculated from the equation

$$T=\frac{P_S}{EA}+To^4\quad 1/4$$

where

E=emissivity
A=surface area of elements
To=ambient temperature
σ=radiation constant

Assuming

E=0.70 (a probable least value)
$\sigma=5.67\times10^{-12}w/cm^2/K^4$
To=600 F

the surface temperature of the sensing element was calculated to be 1,650 F when P_S=7.24 watts, i.e., when $\phi=1\times10^{14}$ neutrons/cm²/second. With the difference P_S-P_B supplied to the balancing element very nearly equal to P_S, it will have, of course, very nearly the same surface temperature.

It can be shown that the difference between the maximum and the surface temperatures of the elements is of the order of a few degrees. Therefore, it is valid to take the calculated surface temperatures as element temperatures.

Magnetic Amplifier Analysis and Application Using Block-Diagram Techniques

ROBERT M. HUBBARD
ASSOCIATE MEMBER AIEE

THE SYNTHESIS of feedback-control systems incorporating magnetic amplifiers has long been hampered by the lack of a really satisfactory way in which to represent the transfer function of the self-saturating magnetic amplifier. Approximations, such as those which represent the magnetic amplifier as being characterized by a volts-per-ampere-turn gain and an *L-R* time constant, are normally unsatisfactory in anything but the simplest system. In addition, there is a need for a single theory of operation which will bridge the existing gap between the analyses of the conventional magnetic amplifier and those of the "half-cycle response" circuits. It is intended that this paper will fill these needs with an analysis utilizing block-diagram techniques. The use of the block diagram is not only compatible with the usual methods of synthesizing feedback-control systems, but also provides an insight into the mechanics of magnetic amplifier operation which heretofore has been lacking.

In this paper, a relatively new philosophy of magnetic amplifier analysis has been adopted. In particular, an effort has been made to examine magnetic amplifier operation without attempting to evaluate specifically those elusive factors associated with dynamic core behavior, rectifier unblocking, signal waveform, etc. It has been found that the effect of all these factors can be lumped into a single circuit parameter, the effective input impedance of the component saturable transformers, which subsequently can be evaluated with varying degrees of accuracy depending on the particular circuit configuration and components used. Indeed, the development of a truly general transfer function precludes the accurate evaluation of these unknown factors prior to drawing the general conclusions. As a result of this method of analysis, the basic simplicity of the magnetic amplifier is shown and the operation of both conventional circuits and half-cycle response circuits is described by a single theory. In addition, the reason for the familiar difficulty in predicting static gain is clarified, and methods for circumventing this problem are suggested.

The techniques employed in this paper were suggested by the work of several other authors.[1–3]

A Philosophy of Magnetic Amplifier Operation

In the past the usual tendency has been to treat magnetic amplifier design as something apart from transformer design. It is proposed that the self-saturating magnetic amplifier is no more than a combination of high-quality transformers applied in a manner which is somewhat unconventional for transformers *per se*. In particular, this device uses saturating transformers, which operate under essentially no-load conditions. Therefore, it is not surprising that most transformer design considerations also apply to magnetic amplifiers.

For example, to obtain easily predictable operation with a transformer operating as a voltage source under no-load conditions, the shunt input impedance (reflected exciting impedance) of the transformer should be negligibly high compared to the signal source impedance. If such is not the case, one is faced with the problem of accurately evaluating the nonlinear exciting impedance. Of course, if the available signal power is of the order of the excitation requirements of the core, the designer is forced in the direction of matching the source impedance to the input impedance. Therefore, in a simple, unloaded transformer the conditions for optimum power transfer and optimum predictability are mutually exclusive. In addition, note that if the source impedance were high compared to the input impedance, one could describe the transformer as a current-controlled device rather than as a voltage-controlled device. Both predictability and sensitivity would be deteriorated in this type of operation. In the final analysis it is apparent that the transformer is inevitably a power-controlled device since the core excitation requirements must be provided by the signal source.

The fact that the foregoing considerations also apply to the self-saturating magnetic amplifier is the basis for the usual difficulty in predicting magnetic amplifier static gain and also accounts for the conflict between the analyses which treat the magnetic amplifier as a voltage-controlled device and those which treat it as a current-controlled device. It will be shown that the class of amplifiers which inherently exhibits half-cycle response is characterized by a driving point impedance which is substantially equal to the input impedance of the component saturable transformers. In contrast, the driving point impedance of the conventional full-wave circuit is reduced by intrinsic, positive voltage feedback to the control winding resistance (under steady-state conditions). Thus the half-cycle response amplifier may be easily treated as a voltage-controlled device while the conventional circuit is ultimately a current-controlled device. Obviously the predictability of the former circuit can be very good while that of the latter is inexorably tied to the accuracy with which the saturable transformer input impedance can be evaluated. This evaluation is complicated by rectifier unblocking which loads the transformer, by the fact that dynamic core behavior is dependent on voltage waveform, etc. Several authors have reported on the results of investigations of these problems; for example, see references 4 and 5.

Development of the Block Diagram

The basic mode of operation of the magnetic amplifier is such that, for purposes of a general transfer function, consideration of increments of time smaller than one-half cycle of the carrier frequency is meaningless. The operation of the magnetic amplifier can, however, be described in terms of half-cycle average values. Attempts to predict the instantaneous relationship between input and output invariably require too many simplifying assumptions to make the resultant analyses generally applicable. Of course, in the case of a particular circuit, such methods are ultimately necessary to the accurate evaluation of those unknown parameters which determine the effective input impedance.

The simplicity of the following analysis

Paper 57-766, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted October 22, 1956; made available for printing May 20, 1957.

ROBERT M. HUBBARD is with the Boeing Airplane Company, Seattle, Wash.

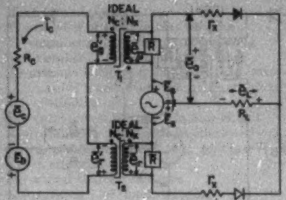
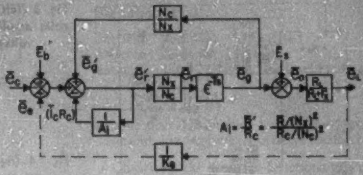


Fig. 1 (left). Full-wave center-tapped magnetic amplifier circuit with component saturable transformers represented in equivalent circuit form

Fig. 2 (right). Block diagram for full-wave magnetic amplifier



is dependent on one important assumption, namely, that during a given half-cycle increment of time, either the component rectifiers do not change their state of conduction or blocking or the effect of any such change is negligible. This assumption is essential to a straightforward, average-value type of analysis since, in view of the dependence of unwanted rectifier unblocking on such things as circuitry, signal voltage waveform, and source impedance level, any attempt to include this factor directly in a general transfer function is doomed to failure.

Fortunately, the resultant general conclusions which are reached with the help of this assumption are not invalidated by the existence of rectifier unblocking. Moreover, for many circuits and operating conditions, the effect of rectifier unblocking on electrical behavior is truly negligible. In particular, this factor usually has little effect on the behavior of center-tapped and bridge circuits operated with rectified a-c signals. In addition, various schemes have been developed to improve magnetic amplifier operation which make the assumption of negligible rectifier unblocking valid. For example, the use of transistors as synchronous rectifiers[6] and the use of the "pulse stretch" circuit[7] accomplish this end. Finally, the use of a square-wave power supply normally assures that the rectifiers will change state only when the supply voltage goes through zero.

Reference to the "center-tapped" circuit of Fig. 1 permits a simple introduction to the block-diagram type of analysis. The two saturable transformers (T_1 and T_2) are represented in equivalent circuit form. The following additional assumptions apply.

1. The core material is of a high remanence type, and the transformer excitation impedance can be represented by an average resistance $\bar{R}$.

2. Transformer leakage reactance is negligible. (For simplicity the input (N_c) winding resistance is lumped with the signal source resistance to produce the total control circuit resistance R_c.)

3. The rectifier reverse impedance is negligibly large compared to $\bar{R}$, and the sum of the load resistance R_L and the output winding plus rectifier forward resistance r_x is negligibly small compared to $\bar{R}$.

4. The supply voltage is sufficient normally to excite the transformers such that the core flux is in saturation at the beginning and end of each cycle of the supply voltage.

5. The volt-time integral absorbed by the output winding N_x during the output (gating) half-cycle equals the volt-time integral applied during the preceding (reset) half-cycle. Therefore, the average gate voltage in the nth half-cycle $[\bar{e}_g(n)]$ equals the average reset voltage in the $(n-1)$th half-cycle $[\bar{e}_r(n-1)]$.

It will be noted that the real departure from the conventional transformer equivalent circuit is in the representation of the core excitation impedance by an equivalent average resistance $\bar{R}$. This factor relates the half-cycle average voltage across the N_x winding to the half-cycle average current through it. This is not an attempt to represent the excitation impedance by equivalent linear elements. The factor $\bar{R}$ is admittedly a nonlinear function of average voltage and its magnitude is dependent on the waveforms involved. In fact, all of the unknown core parameters are lumped into $\bar{R}$. The fact that high-remanence cores produce essentially "in-phase" exciting currents is the only justification for designating the excitation impedance as a resistance. The accurate evaluation of $\bar{R}$ for a particular circuit can be a formidable problem and is rightfully the subject of a separate investigation. Suffice it to say that simple core-test methods permit the evaluation of $\bar{R}$ to within a factor of two regardless of the waveforms of voltage and current involved. It should be noted in passing that the bulk of the nonlinearity normally found in the magnetic amplifier static transfer characteristic can be attributed to the nonlinear nature of the exciting impedance $\bar{R}$.

The following description of the operation of the simple magnetic amplifier of Fig. 1 based on the equivalent block diagram of Fig. 2 indicates the straightforward relationship between the two. An analytical development of the block diagram is presented in the appendix.

During the indicated half-cycle of the supply voltage the upper rectifier is conducting and the lower one is blocking. Transformer T_1 is therefore gating and transformer T_2 is resetting. The average reset voltage ($\bar{e}_r'$) applied to the control winding of T_2 is the sum of three voltages. These are the net signal ($\bar{E}_b-\bar{e}_c$), the intrinsic, delayed, positive-feedback voltage ($\bar{e}_g'$) produced by the series-connected control windings, and the intrinsic negative feedback voltage ($\bar{i}_cR_c$) due to the finite excitation requirements of the core. The reset voltage referred to the load side ($\bar{e}_r$) in the given half-cycle is equal to the gate voltage ($\bar{e}_g$) effective on the subsequent half-cycle; thus the half-cycle time delay, (ϵ^{-Ts}), is introduced using Laplace notation. The average output voltage $\bar{e}_o$ is the difference between the average supply voltage $\bar{E}_s$ and the average gate voltage $\bar{e}_g$. The load voltage $\bar{e}_L$ is merely that portion of the output which appears across the load resistor R_L.

A steady state will be achieved when the gating voltage in one half-cycle equals the gating voltage in the subsequent half-cycle. Therefore, in a given half-cycle under steady-state conditions, the gating voltage $\bar{e}_g'$ must equal the reset voltage $\bar{e}_r'$. From the block-diagram it can be seen that a necessary condition for a steady-state is that the sum of the net input voltage ($E_b-\bar{e}_c$) and the intrinsic negative feedback voltage ($\bar{i}_cR_c$) be zero. This is, of course, exactly what one expects. The symmetrical nature of the circuit of Fig. 1 makes the block diagram valid for both half-cycles of the supply voltage.

Observations Based on the Block Diagram

It has been pointed out that in the interest of predictability the driving point impedance of the amplifier should be high compared to the signal source impedance. Obviously this is an impossible condition to achieve in the foregoing circuit since the d-c driving point impedance is reduced to zero by the intrinsic positive feedback, if the control winding resistance

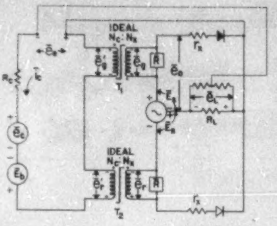
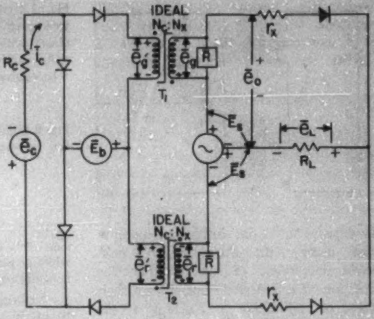


Fig. 3 (left). Full-wave magnetic amplifier with external voltage feedback

Fig. 4 (right). Ramey-type magnetic amplifier characterized by half-cycle response



is lumped with the source impedance. Even the more proper inclusion of the the control winding resistance in the driving point impedance would not serve to ease the situation. Under these conditions the magnetic amplifier is most assuredly a current-controlled device and the prediction of the static gain is dependent on the evaluation of the intrinsic amplification factor $(A_i = \bar{R}'/R_c)$. A real improvement in predictability can be achieved by applying external, negative voltage feedback to raise the driving point impedance to a level substantially higher than the control circuit resistance R_c. This additional delayed feedback may simply be added in series with the input as indicated by the dotted lines of Fig. 2, and the modified circuit of Fig. 3. The feedback factor $(1/K_e)$ is merely the ratio of feedback voltage $\bar{e}_e$ to load voltage $\bar{e}_L$. The resistive feedback network will of course modify the effective load resistance and control circuit resistance.

In those classes of circuits in which the intrinsic positive feedback $(\bar{e}_g')$ is either absent (e.g., half-wave circuits) or has been made ineffective, the predictability can be improved simply by increasing the intrinsic amplification factor A_i. One of the latter circuits is the Ramey-type circuit of Fig. 4 in which the intrinsic positive feedback is blocked by rectifiers.[9] The block diagram appears in Fig. 5.

It is of interest to note that the "time constant" normally associated with full-wave magnetic amplifiers does not appear in the proposed block diagram. It will be shown that this lag is a direct result of the intrinsic positive feedback produced by the series-connected control windings. Without this positive feedback the amplifier would merely have a half-cycle delay as indicated by the ϵ^{-Ts} term. Since no intrinsic positive feedback is possible in a simple half-wave circuit, the block diagram of Fig. 5 is valid for this class of circuits as well as for the circuit of Fig. 4. It will be shown that, in the case of the full-wave circuit of Fig. 1, external negative feedback can be used to nullify completely the intrinsic positive feedback in order to produce "half-cycle response" in this circuit also. Obviously, care must be observed in applying the block diagram to a simple half-wave circuit since an input signal can be accepted only during one half-cycle and an output obtained only during the subsequent single half-cycle.

The use of various external feedback circuits allows one to feed back either a current or a voltage proportional to either the output voltage or current. Thus the effective input impedance and output impedance may be easily modified. Of course, amplifiers in which the input and output quantities differ, such as d-c input–a-c output, present problems in deriving the feedback signal. The subject of input impedance and output impedance is treated in somewhat more detail later.

If the core material were idealized so as to have no excitation requirements (i.e., $\bar{R} \to \infty$), there would be no intrinsic negative feedback. The resultant amplifier would act as an integrator with an inherent time delay. Although such a device cannot be used as a proportional amplifier *per se*, this type of operation cannot be considered as instability in the usual sense of the term. This condition is roughly equivalent to that obtained by applying sufficient extrinsic positive feedback to nullify the intrinsic negative feedback and is referred to as "snap action." It is a type of operation which is utilized in the "self-balancing potentiometer" type of circuit, which will be briefly discussed subsequently.[9]

Nomenclature

A = amplification factor
A_b = bias circuit amplification factor (multiple reset winding circuit)
A_c = control circuit amplification factor (multiple reset winding circuit)
A_e = extrinsic amplification factor
A_i = intrinsic amplification factor
A_o = generalized amplification factor
$\bar{E}_b$ = average steady-state bias voltage
$\bar{e}_c$ = half-cycle average control voltage
$\bar{E}_c$ = average steady-state control voltage
$\bar{e}_e$ = half-cycle average external feedback voltage
$\bar{e}_g$ = half-cycle average gate voltage across load winding
$\bar{e}_g'$ = half-cycle average gate voltage referred to control winding
$\bar{e}_{gb}'$ = half-cycle average gate voltage referred to bias winding (multiple reset winding circuit)
$\bar{e}_{gc}'$ = half-cycle average gate voltage referred to control winding (multiple reset winding circuit)
$\bar{e}_L$ = half-cycle average load voltage (across R_L)
$\bar{e}_o$ = half-cycle average output voltage (across $R_L + r_x$)
$\bar{E}_o$ = average steady-state output voltage
$\bar{e}_r$ = half-cycle average reset voltage across load winding
$\bar{e}_r'$ = half-cycle average reset voltage referred to control winding
$\bar{E}_s$ = average rectified supply voltage
G_p = power gain
$G(s)$ = frequency variant portion of transfer function, $K\,G(s)$
$\bar{i}_b$ = half-cycle average current in bias winding
$\bar{i}_c$ = half-cycle average current in control winding
K = static voltage gain associated with transfer function, $K\,G(s)$
K_e = reciprocal of external feedback gain (ratio of load voltage to feedback voltage)

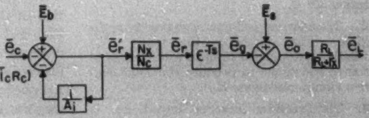


Fig. 5. Block diagram of half-cycle response magnetic amplifier

M = figure of merit
N_b = number of turns in bias winding
N_c = number of turns in control winding
N_z = number of turns in load winding
$\bar{R}$ = average excitation impedance of saturable transformer referred to the load winding
R_b = total series resistance in bias circuit
R_c = total series resistance in control circuit
r_c = resistance of control (N_c) windings
R_L = load resistance
R_s = signal source resistance
r_z = resistance of load winding plus the forward resistance of the rectifier
s = complex frequency variable
T = time for one half-cycle of supply (ω_c)
τ_a = amplifier 63% response time
ω_c = supply frequency in radians per second
ω_0 = "break frequency" or frequency for 45-degree phase shift
Z_i = amplifier driving point impedance
Z_0 = amplifier output impedance
η_c = control circuit efficiency

Magnetic Amplifier Transfer Functions

The transfer function for the basic amplifier of Fig. 1 can be written in operational form directly from the block diagram as follows:

$$-\left[-\bar{e}_c+E_b+\left(\frac{N_c}{N_z}\right)\bar{e}_g\right]\times\left[\frac{1}{1+\dfrac{1}{A_i}}\right]\left[\frac{N_z}{N_c}\right]\epsilon^{-Ts}+E_s=\bar{e}_o$$

By substituting the proper bias,

$$E_b=\left[\frac{N_c}{N_z}\right]\left[\frac{1}{A_i}\right]E_s$$

and

$$E_s=\bar{e}_g+\bar{e}_o$$

the foregoing equation becomes

$$\left[\bar{e}_c+\left(\frac{N_c}{N_z}\right)\bar{e}_o\right]\left[\frac{1}{1+\dfrac{1}{A_i}}\right]\left[\frac{N_z}{N_c}\right]\epsilon^{-Ts}=\bar{e}_o$$

This may be simplified to

$$\frac{\bar{e}_o}{e_c}=\frac{N_z}{N_c}\left[\frac{\epsilon^{-Ts}}{\left(1+\dfrac{1}{A_i}\right)-\epsilon^{-Ts}}\right] \tag{1}$$

The addition of delayed negative feedback as indicated in Fig. 3 produces the following transfer function:

$$\frac{\bar{e}_L}{\bar{e}_c}=\left[\frac{R_L}{R_L+r_z}\right]\left[\frac{N_z}{N_c}\right]\times\left[\frac{\epsilon^{-Ts}}{\left(1+\dfrac{1}{A_i}\right)-\left(1-\dfrac{1}{A_e}\right)\epsilon^{-Ts}}\right] \tag{2}$$

in which

$$A_e=\left(\frac{N_f}{N_z}\right)\left(\frac{\bar{e}_o}{\bar{e}_e}\right)=\left(\frac{N_c}{N_z}\right)\left(\frac{R_L+r_z}{R_L}\right)K_e$$

The foregoing is better put in the form of equation 3, as follows:

$$\frac{\bar{e}_o}{\bar{e}_c}=\frac{N_z}{N_c}\left[\frac{\epsilon^{-Ts}}{\left(1+\dfrac{1}{A_i}\right)-\left(1-\dfrac{1}{A_e}\right)\epsilon^{-Ts}}\right] \tag{3}$$

If $A_i\rightarrow\infty$ only the delayed feedback is effective and

$$\frac{\bar{e}_o}{\bar{e}_c}=\frac{N_z}{N_c}\left[\frac{\epsilon^{-Ts}}{1-\left(1-\dfrac{1}{A_e}\right)\epsilon^{-Ts}}\right] \tag{4}$$

The foregoing expressions apply for negative feedback. If positive feedback is used, the signs of the appropriate amplification factors (A) become negative.

The three basic expressions in equations 1, 3, and 4 may be rewritten in terms of static gain K and the frequency dependent variable $G(s)$ as follows:

Intrinsic feedback:

$$K\,G(s)=\left[(A_i)\frac{N_z}{N_c}\right]\left[\frac{\epsilon^{-Ts}}{(A_i+1)-(A_i)\epsilon^{-Ts}}\right] \tag{5}$$

Intrinsic plus delayed negative feedback:

$$K\,G(s)=\left[\left(\frac{A_iA_e}{A_i+A_e}\right)\left(\frac{N_z}{N_c}\right)\right]\left[\left(\frac{A_i+A_e}{A_iA_e}\right)\times\frac{\epsilon^{-Ts}}{\left(1+\dfrac{1}{A_i}\right)-\left(1-\dfrac{1}{A_e}\right)\epsilon^{-Ts}}\right] \tag{6}$$

Delayed negative feedback only:

$$K\,G(s)=\left[(A_e)\left(\frac{N_z}{N_c}\right)\right]\times\left[\frac{\epsilon^{-Ts}}{(A_e)-(A_e-1)\epsilon^{-Ts}}\right] \tag{7}$$

It has been found that, in the interest of simplicity, the frequency-dependent part of the transfer functions for these and a variety of other configurations can be reduced to the basic form:

$$G(s)=\left[\frac{\epsilon^{-Ts}}{(A_0)-(A_0-1)\epsilon^{-Ts}}\right] \tag{8}$$

The evaluation of the generalized amplification factor A_0 is dependent on the particular circuit under consideration. For example, the frequency-dependent portions of the three transfer functions may be generalized by making the following substitutions:

Intrinsic feedback: $A_0=A_i+1$
Delayed feedback: $A_0=A_e$
Intrinsic plus delayed feedback: $A_0=\dfrac{A_i+1}{(A_i/A_e)+1}$

Note that A is positive for negative feedback and A is negative for positive feedback.

Obviously the transfer functions given are in a form which is not readily adaptable to the synthesis of feedback control systems. Since there is not a finite number of roots in the s-plane associated with equation 8, the magnetic amplifier does not lend itself to root-locus methods of synthesis. Fortunately it does lend itself to the use of the Bode diagram.

Equation 8 rewritten as a function of steady-state angular frequency, becomes

$$G(j\omega)=\left[\frac{1}{1-2A_0(1-A_0)\left(1-\cos\pi\dfrac{\omega}{\omega_c}\right)}\right]^{1/2}\angle\phi \tag{9}$$

where

$$\phi=-\tan^{-1}\left[\frac{\sin\pi\dfrac{\omega}{\omega_c}}{\cos\pi\dfrac{\omega}{\omega_c}-\left(1-\dfrac{1}{A_0}\right)}\right] \tag{10}$$

and ω_c is the carrier (supply voltage) angular frequency.

If ω_0 is the frequency at which $\phi=45$ degrees,

$$\frac{\omega_0}{\omega_c}=\frac{1}{2\pi}\sin^{-1}\left[\frac{2A_0-1}{A_0^2}\right] \tag{11}$$

The substitution of equation 11 into equations 9 and 10 yields

$$(j\omega)=\left[\frac{1}{1-[2A_0(1-A_0)]\left[1-\cos\left\{\dfrac{\omega}{2\omega_0}\sin^{-1}\left(\dfrac{2A_0-1}{A_0^2}\right)\right\}\right]}\right]^{1/2}\angle\phi \tag{12}$$

and

$$\phi=-\tan^{-1}\left[\frac{\sin\left[\dfrac{\omega}{2\omega_0}\sin^{-1}\left(\dfrac{2A_0-1}{A_0^2}\right)\right]}{\cos\left[\dfrac{\omega}{2\omega_0}\sin^{-1}\left(\dfrac{2A_0-1}{A_0^2}\right)\right]-\left[1-\dfrac{1}{A_0}\right]}\right] \tag{13}$$

Equations 12 and 13 are plotted in Fig. 6 and equation 11 is plotted in Fig. 7. Note that in one limit ($A_0=1$) the magnetic amplifier is comparable to an ideal delay line while in the other limit ($A_0\rightarrow\infty$) it approaches the operation to be expected from a simple first-order system characterized by a single time constant. This is even more clearly emphasized by the polar co-ordinate plot of Fig. 8. Note that for values of A_0 greater than 10, $G(j\omega)$ is approximately that of a linear system with a single time constant. However, for finite values of A_0, the

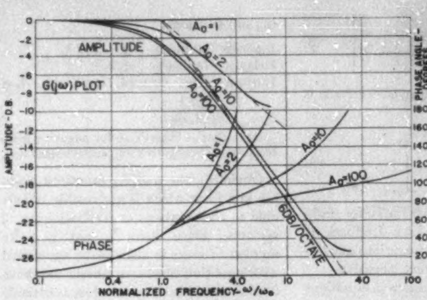


Fig. 6. Phase-amplitude plot of the transfer function K G(s)/K for the generalized, self-saturating magnetic amplifier

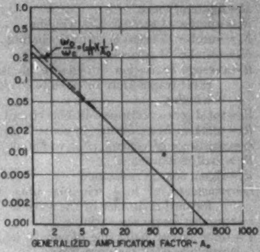


Fig. 7. Normalized break frequency versus generalized amplification factor, A₀

phase-shift always reaches 180 degrees in the limit ($\omega=\omega_c$). Of course, to eliminate the phase detector action of the magnetic amplifier it is assumed that the signal frequency is in phase with the carrier frequency in this limit. The foregoing clarifies the reasons for assuming the magnetic amplifier to be characterized by a single time constant and permits the evaluation of the error involved in such an assumption.

If $A_o \geq 10$ the relationship between the break frequency ω_0 and the generalized amplification factor A_o can be expressed with good approximation (see Fig. 7 and the appendix) as

$$\frac{\omega_0}{\omega_c}=\left[\frac{1}{\pi}\right]\left[\frac{1}{A_o}\right] \tag{14}$$

Therefore an equivalent time constant τ_a for the amplifier response to a step function can be written in the following manner:

$$\tau_a=\frac{1}{\omega_0}=\left(\frac{\pi}{\omega_c}\right)(A_o)$$

If further, one lets $\pi/\omega_c=T$, which is the time for a half-cycle of the carrier voltage, then

$$\tau_a=A_oT,\ [A_o\geq 10] \tag{15}$$

It will be shown subsequently that τ_a is better referred to as the "63% response time" than as the "time constant." As such, τ_a can be utilized without restrictions on the value of A_o.

Step-Function Response

A consistent half-cycle average type of analysis precludes the representation of the magnetic amplifier input or output as anything but constant at the average value over each half-cycle. In Fig. 9 is shown the normalized response to a a step function which was obtained by a numerical evaluation of equation 16. Since

$$\frac{\bar{e}_o}{\bar{e}_c}=KG(s)$$

$$K=\frac{E_o}{E_c}$$

and

$$\bar{e}_c=\frac{E_c}{s}$$

it follows that

$$\frac{\bar{e}_o}{E_o}=\frac{1}{s}[G(s)] \tag{16}$$

It is assumed that the input step occurs at the beginning of a half-cycle increment of time. This assumption is of course essential in order to obtain a step change in the average value of the control voltage. A step function of voltage introduced at any other time would produce an input whose average value is different during the second half-cycle than during the first and thereby produce a double transient in the output. The smoothed response is a half-cycle delay followed by an exponential rise. It is of particular interest to note that, in every case, at least 63% response is achieved at time, $t=A_oT$. For large values of A_o, where the initial half-cycle delay becomes negligible, the response approaches a true exponential. The fact that the response is definitely not exponential for small values of A_o indicates the reason for referring to τ_a as the 63% response time rather than the time constant. The response for $A_o=1$ corresponds to half-cycle response operation.

The response to a unit step voltage, Fig. 10, for the circuit of Fig. 3 was obtained in a manner similar to that used in the foregoing. For simplicity the intrinsic feedback has been assumed negligi-

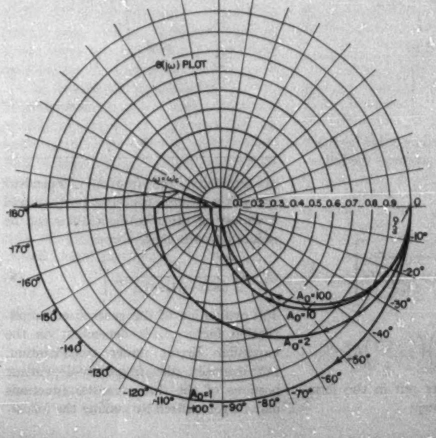


Fig. 8. Polar coordinate plot of the transfer function K G(s)/K for the generalized, self-saturating magnetic amplifier

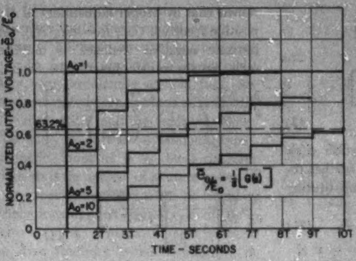


Fig. 9 (left). Normalized voltage response of the generalized amplifier to a step-function driving voltage

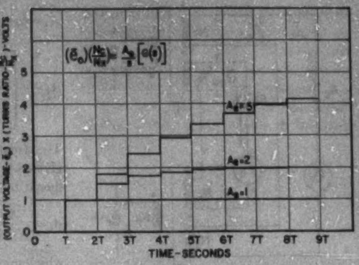


Fig. 10 (right). Voltage response to a unit step-function driving voltage for the circuit of Fig. 3 ($A_i \to \infty$)

ble compared to the extrinsic feedback. Thus, since

$$\frac{\bar{e}_o}{\bar{e}_c} = K\,G(s)$$

$$K = A_e\left(\frac{N_x}{N_c}\right)$$

and

$$\bar{e}_c = \frac{1}{s}$$

then

$$\bar{e}_o\left(\frac{N_c}{N_x}\right) = \frac{A_e}{s}[G(s)] \tag{17}$$

Note that for a given input, the output during the second half-cycle is independent of A_e and thus independent of the ultimate steady-state output. Obviously this must be the case since the effect of any feedback is delayed until the third half-cycle. This plot indicates the increase in output that can be achieved by providing for longer response times.

Obtaining Fast Response With Magnetic Amplifiers

There are three basic methods for obtaining fast response in ordinary magnetic amplifier circuits. First, a half-wave-type circuit or a Ramey-type circuit which is characterized by half-cycle response may be used. The block diagram for such a circuit appears in Fig. 5. The transfer function is simply as follows:

$$\frac{\bar{e}_o}{\bar{e}_c} = \left[\frac{N_x}{N_c}\right]\left[\frac{1}{1+\frac{1}{A_i}}\right]\epsilon^{-\tau s} \tag{18}$$

The second method is to add "forcing resistance" to the control circuit of a simple full-wave circuit. This increase in R_c reduces the voltage gain and the intrinsic amplification factor A_i. However, one reaches a practical limit when the source resistance R_c equals the input resistance $\bar{R}'$ (i.e., $A_i = 1$). The resultant response for $A_i = 1$ can be determined by noting that

$$A_e = A_i + 1 = 2$$

and

$$\tau_R = A_e T = 2$$

Thus a practical limit for the simple full-wave circuit is a response time of two half-cycles of the carrier frequency. Note that to obtain true half-cycle response the total source resistance R_c must approach infinity. This limiting condition is of questionable practical value but does provide the necessary simplification for some types of theoretical analyses. Since the assumption of negligible rectifier unblocking can be grossly in error for conditions where $A_i << 1$, the techniques developed in this paper can give questionable results in this region of operation.

The third method of response improvement involves adding extrinsic, delayed negative feedback to a full-wave-type circuit. Scorgie and House have described the limiting case in which the resultant response-time is one half-cycle.[1,10] Consider the transfer function, equation 3, which applies to the circuit of Fig. 3. If the extrinsic amplification factor A_e is made equal to one, equation 3 reduces to

$$\frac{\bar{e}_o}{\bar{e}_c} = \left[\frac{N_x}{N_c}\right]\left[\frac{1}{1+\frac{1}{A_i}}\right]\epsilon^{-\tau s} \tag{19}$$

Thus half-cycle response is achieved independent of the value of A_i. The factor A_i only affects the gain under these conditions. The foregoing expression is also the transfer function for the Ramey-type crcuit of Fig. 4. In fact, the addition of the appropriate amount of extrinsic feedback ($A_e = 1$), makes the circuit of Fig. 3 electrically equivalent to the circuit of Fig. 4. This can be seen by examining the block diagram of Fig. 11 which is valid for the full-wave circuit in which $A_e = 1$ (i.e., $1/K_e = N_c/N_x$) and $R_L >> r_x$. Note that the sum of the intrinsic feedback $(\bar{e}_g)(N_c/N_x)$ and the extrinsic feedback $(\bar{e}_o)$ (N_c/N_x) equals a constant $(\bar{E}_s)(N_c/N_x)$. The block diagram of Fig. 11 is therefore equivalent to the block diagram of the half-wave-type circuit (Fig. 5).

The full-wave circuit with external negative feedback has the advantage of versatility over those circuits with intrinsic half-cycle response in that the gain and response time can be varied over a wide range to suit the particular application.

Amplifiers With Multiple Reset Windings

Normally a magnetic amplifier has at least two separate reset windings, one for the signal and the other for bias. The block-diagram approach reveals some interesting facts about such a circuit. The reset circuit of a 2-winding magnetic amplifier is represented in Fig. 12 and the appropriate block diagram appears in Fig. 13. A simplified block diagram based on the application of the Superposition Theorem appears in Fig. 14.

The transfer function for the foregoing circuit is as follows:

$$\frac{\bar{e}_o}{\bar{e}_c} = \frac{N_x}{N_c}\left[\frac{\epsilon^{-\tau s}}{\left[1+\frac{1}{A_c}\right] - \left[1+\frac{\left(1+\frac{1}{A_c}\right)}{\left(1+\frac{1}{A_b}\right)}\right]\epsilon^{-\tau s}}\right] \tag{20}$$

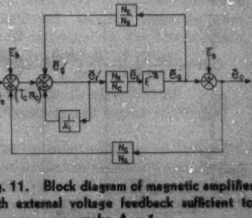


Fig. 11. Block diagram of magnetic amplifier with external voltage feedback sufficient to make $A_e = 1$

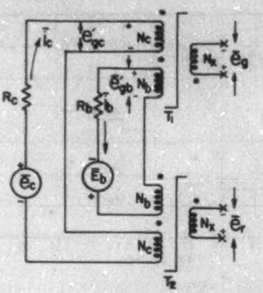


Fig. 12. Modified reset circuit for the full-wave center-tapped magnetic amplifier with two reset windings

In the foregoing

$$A_c=\frac{[(N_c)^2/R_c]}{[(N_x)^2/\bar{R}]+[(N_b)^2/R_b]}$$

and

$$A_b=\frac{[(N_b)^2/R_b]}{[(N_x)^2/\bar{R}]+[(N_c)^2/R_c]}$$

Equation 20 is in the general form of the transfer function for a circuit with intrinsic negative feedback and delayed positive feedback. The positive feedback is introduced by the series connection of the bias windings in the same manner that the intrinsic positive feedback is introduced into the control winding.

Equation 20 may be put in the more general form of equation 3 by making the indicated substitutions.

$$\frac{\bar{e}_o}{\bar{e}_c}=\frac{N_x}{N_c}\left[\frac{\epsilon^{-Ts}}{\left(1+\frac{1}{A_i}\right)-\left(1-\frac{1}{A_e}\right)\epsilon^{-Ts}}\right]$$

$$A_i=A_c=\frac{[(N_c)^2/R_c]}{[(N_x)^2/\bar{R}]+[(N_b)^2/R_b]}$$

$$A_e=-\left[\frac{1+\frac{1}{A_b}}{1+\frac{1}{A_c}}\right]=-\left[\frac{(N_c)^2/R_c}{(N_b)^2/R_b}\right] \tag{3}$$

The frequency dependent portion $G(s)$, given by equation 8, may be further reduced to the general case by substituting

$$A_o=\frac{A_i+1}{(A_i/A_e)+1} \tag{21}$$

which may be rewritten in terms of circuit resistances as

$$A_o=\left[\frac{[(N_c)^2/R_c]+[(N_b)^2/R_b]}{[(N_x)^2/\bar{R}]}+1\right] \tag{22}$$

The foregoing expression infers an effect of circuit resistance on response time which is entirely consistent with presently accepted theories.

Amplifiers With Multiple External Feedback Loops

The addition of sufficient external positive feedback to nullify the intrinsic negative feedback in a conventional full-wave magnetic amplifier produces a circuit with infinite static gain. The subsequent application of external negative feedback produces a stable amplifier with an exactly predictable static transfer characteristic. This principle has been the subject of several papers.[9,11,12] Normally one is concerned with the complete transfer function of an amplifier rather than with merely the static portion. The block-diagram analysis provides this complete transfer function and perhaps helps clarify the operation of this type of circuit.

Consider the application of positive feedback to the bias (N_b) winding of the circuit of Fig. 12 (in which $R_L >> r_x$) as indicated by the lower dotted line on the block diagram (Fig. 14). It can be shown that by proper adjustment of the positive feedback (i.e., $R_b=\bar{R}N_b/N_x$), the intrinsic negative feedback is nullified such that the transfer function becomes

$$\frac{\bar{e}_o}{\bar{e}_c}=\frac{N_x}{N_c}\left[\frac{\epsilon^{-Ts}}{\left(1+\frac{1}{A_c}\right)-\left(1+\frac{1}{A_c}\right)\epsilon^{-Ts}}\right] \tag{23}$$

or

$$\frac{\bar{e}_o}{\bar{e}_c}=\frac{N_x}{N_c}\left[\frac{1}{1+\frac{1}{A_c}}\right]\left[\frac{\epsilon^{-Ts}}{1-\epsilon^{-Ts}}\right] \tag{24}$$

Note that

$$\text{Limit}\ \frac{\bar{e}_o}{\bar{e}_{c\,s\to 0}}=\infty$$

Thus the static gain becomes infinite. The step-function response of this circuit is that of an integrator with an inherent half-cycle time delay. The average output is a linear function of the number of increments of time, (nT).

The addition of a stabilizing external negative feedback produces a usable proportional amplifier. One method of applying this feedback, indicated by the upper dotted lines of Fig. 14, is by feeding back a fraction of the output voltage to the control winding directly in series with the signal. The closed-loop transfer function then becomes

$$\frac{\bar{e}_o}{\bar{e}_c}=K\,G(s)=\left[(A_e)\left(\frac{N_x}{N_c}\right)\right]\times$$
$$\left[\frac{\epsilon^{-Ts}}{\left[A_e\left(1+\frac{1}{A_c}\right)\right]-\left[A_e\left(1+\frac{1}{A_c}\right)-1\right]\epsilon^{-Ts}}\right] \tag{25}$$

and therefore

$$\text{Limit}\ \frac{\bar{e}_o}{\bar{e}_{c\,s\to 0}}=(A_e)\left(\frac{N_x}{N_c}\right)=K_e$$

Note that although the static gain is independent of the basic amplifier characteristics, the frequency-dependent por-

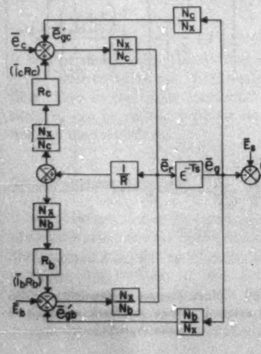


Fig. 13 (left). Block diagram of the magnetic amplifier with two reset windings

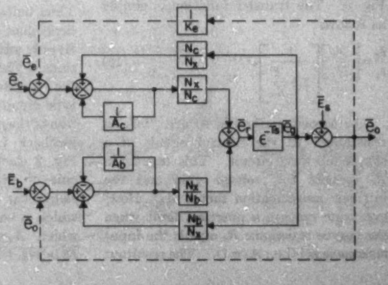


Fig. 14 (right). Simplified block diagram of the magnetic amplifier with two reset windings

tion of the transfer function is directly dependent on the basic amplifier gain.

The techniques presented in this paper may be readily applied to considerably more complex circuits and, in fact, are most valuable in the analysis of magnetic amplifiers with multiple feedback loops. The analysis of the "operational-type" magnetic amplifier[11] by block-diagram methods is the subject of a future paper.

In general, the feedback to be found in magnetic amplifier circuitry is of three general types. First, it may be intrinsic negative feedback which is effective in reducing the net reset voltage; second, it may be delayed feedback from output to input; third, it may be feedback in the load circuit due to mixing two outputs into a common load. This last feedback is of prime importance in the analysis of differential output circuits.

Gain Bandwidth Product

An important parameter associated with amplifiers in general is the gain-bandwidth product. If one defines the magnetic amplifier bandwidth as corresponding to the break frequency ω_0, simple expressions for the gain bandwidth $(K\cdot\omega_0)$ can be derived. For the amplifier with both intrinsic negative feedback and external (delayed) feedback:

$$K=\frac{E_0}{E_c}=\left[\frac{A_iA_e}{A_i+A_e}\right]\left[\frac{N_x}{N_c}\right]$$

and

$$\omega_0=\left(\frac{\omega_c}{\pi}\right)\left(\frac{1}{A_0}\right), \quad [A_0\geq 10]$$

Substituting $A_0=A_i+1/(A_i/A_e)+1$, and simplifying produces the following:

$$(K\cdot\omega_0)=\left[\frac{N_x}{N_c}\right]\left[\frac{1}{1+\frac{1}{A_i}}\right]\left[\frac{\omega_c}{\pi}\right], \begin{cases}A_0\geq 10 \text{ or}\\ A_i\geq 9\end{cases} \tag{26}$$

Note that $(K\cdot\omega_0)$ is independent of the extrinsic feedback, a fact which has been emphasized by Johannessen.[3] Note also that in the range of operation where the magnetic amplifier may be treated as a simple first-order system, the gain-bandwidth product is approximately a constant independent of the intrinsic feedback, i. e.

$$(K\cdot\omega_0)\simeq\left(\frac{N_x}{N_c}\right)\left(\frac{\omega_c}{\pi}\right), [A_0\geq 10] \tag{27}$$

If, instead of the gain-bandwidth product, one prefers to evaluate the ratio of gain to response time, the limitation imposed on the value of A_0 to obtain a simple expression is eliminated. Thus,

$$\frac{K}{\tau_a}=\frac{K}{A_0T}$$

and

$$\frac{K}{\tau_a}=\left[\frac{N_x}{N_c}\right]\left[\frac{1}{1+\frac{1}{A_i}}\right]\left[\frac{1}{T}\right] \tag{28}$$

or

$$\frac{K}{\tau_a}=\left[\frac{N_x}{N_c}\right]\left[\frac{1}{1+\frac{1}{A_i}}\right]\left[\frac{\omega_c}{\pi}\right] \tag{29}$$

It is of interest to examine the condition for maximum K/τ_a. Since

$$A_i=\left(\frac{\hat{R}'}{R_c}-\frac{\hat{R}}{R_c}\right)\left(\frac{N_c}{N_x}\right)^2$$

it follows that

$$\frac{K}{\tau_a}=\frac{(\omega_c/\pi)}{\left(\frac{N_c}{N_x}\right)+\frac{(R_c/\hat{R})}{(N_c/N_x)}}$$

and

$$\frac{\partial(K/\tau_a)}{\partial(N_c/N_x)}=\left[\frac{(R_c/\hat{R})}{(N_c/N_x)^2}-1\right]\frac{\omega_c}{\pi} \tag{30}$$

By setting the derivative equal to zero, the condition for maximum K/τ_a with (N_c/N_x) as a variable becomes

$$R_c=\hat{R}\left[\frac{N_c}{N_x}\right]^2 \tag{31}$$

Driving Point Impedance and Output Impedance

A rigorous examination of the subject impedances demands that the control winding resistance r_c be separated from the source resistance R_s so that the total resistance in the control circuit R_c need not be mentioned. It must be assumed that any resistances introduced by feedback networks are lumped into the source resistance R_s.

The driving point impedance of the Ramey-type circuit of Fig. 4 can be determined by inspection to be the sum of the control winding resistance r_c and the saturable transformer input impedance $\hat{R}'$. The same is true of half-wave-type circuits during the control half-cycle.

The intrinsic positive voltage feedback of the conventional full-wave circuit can be shown (see appendix) to reduce the driving point impedance from $(r_c+\hat{R}')$ to

$$Z_i=r_c+\hat{R}'(1-\epsilon^{-\tau_e}) \tag{32}$$

and $\quad \operatorname{Limit}_{s\to\infty} Z_i=r_c$

The point has already been made that the inevitably low driving point impedance of the conventional full-wave magnetic amplifier makes it a current-controlled device under steady-state conditions. However, during transients, this impedance is higher and the voltage-controlled characteristics are apparent.

The more general case for the circuit of Fig. 3 with external voltage feedback is given by

$$Z_i=r_c+\hat{R}'\left[1-\left(1-\frac{1}{A_e}\right)\epsilon^{-\tau_e}\right] \tag{33}$$

and $\quad \operatorname{Limit}_{s\to\infty} Z_i=r_c+\hat{R}'/A_e$

Note that if the negative voltage feedback is sufficient to give half-cycle response $(A_e=1)$ the driving point impedance becomes

$$Z_i=(r_c+\hat{R}'), \qquad [A_e=1] \tag{34}$$

This, of course, merely emphasizes the equivalence of this circuit to those circuits which intrinsicly exhibit half-cycle response. If the external feedback is removed $(A_e\to\infty)$ the special case of the simple circuit obtains and the driving point impedance is given by equation 32.

The output impedance of any magnetic amplifier without external feedback is obviously equal to r_x, the sum of the load winding and rectifier forward resistances. However, in the circuit of Fig. 3, for example, the internal impedance r_x is inside the negative feedback loop and can therefore be radically reduced. The output impedance can be found from equation 2 to be

$$Z_0=r_x\left[\frac{\left(1+\frac{1}{A_i}\right)-\epsilon^{-\tau_e}}{\left(1+\frac{1}{A_i}\right)-\left(1-\frac{1}{A_e}\right)\epsilon^{-\tau_e}}\right] \tag{35}$$

If $A_e=1$,

$$Z_0=r_x\left[1-\left(\frac{A_i}{1+A_i}\right)\epsilon^{-\tau_e}\right], [A_e=1] \tag{36}$$

and $\quad \operatorname{Limit}_{s\to 0} Z_0=r_x\left[\frac{1}{1+\frac{1}{A_i}}\right]$

Thus the equivalence of the Ramey-type circuit and the conventional circuit with 100% external voltage feedback $(A_e=1)$ breaks down when the output impedance is considered. In many cases the lower output impedance of the latter circuit represents a distinct advantage.

Power Gain, Control Circuit Efficiency, and Figure of Merit

The rigorous evaluation of the power gain of a magnetic amplifier is complicated by the variations possible in the form factors of both the input and the output waveforms. The development of a gen-

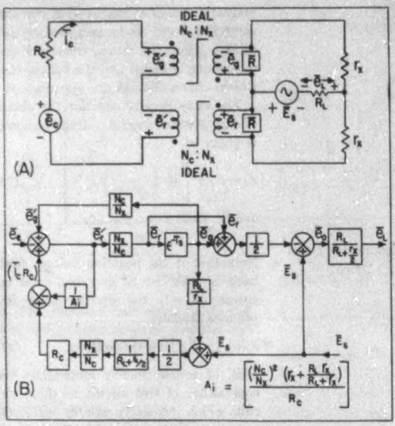

IDEAL
$N_c : N_x$
I_c
R_c
$\bar{e}_g'$
$\bar{e}_r'$
$\bar{e}_c$
R
$\bar{e}_g$
$\bar{e}_r$
$\bar{e}_L$
r_x
R_L
E_s
(A)
(B)
$\frac{N_c}{N_x}$
$\frac{N_x}{N_c}$
$\epsilon^{-\tau_s}$
$\frac{1}{2}$
$\frac{R_L}{R_L+\frac{r_x}{2}}$
$(\bar{I}_c R_c)$
$\frac{1}{A_i}$
$\frac{R_L}{r_x}$
R_c
$\frac{1}{R_L+r_x/2}$
$A_i = \left[\frac{\left(\frac{N_c}{N_x}\right)^2 \left(r_x + \frac{R_L r_x}{R_L+r_x}\right)}{R_c}\right]$

Fig. 15. A—Parallel saturable reactor. B—Block diagram for parallel saturable reactor

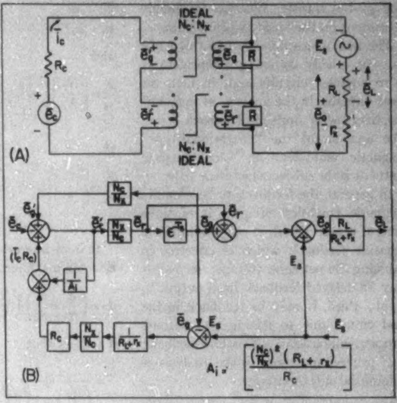

IDEAL
$N_c : N_x$
I_c
R_c
$\bar{e}_c$
$\bar{e}_g'$
$\bar{e}_r'$
$\bar{e}_g$
$\bar{e}_r$
R
E_s
R_L
$\bar{e}_L$
$\bar{e}_o$
r_x
(A)
(B)
$\frac{N_c}{N_x}$
$\frac{N_x}{N_c}$
$\epsilon^{-\tau_s}$
$\frac{R_L}{R_L+r_x}$
$(\bar{I}_c R_c)$
$\frac{1}{A_i}$
R_c
$\frac{1}{R_L+r_x}$
$A_i = \left[\frac{\left(\frac{N_c}{N_x}\right)^2 (R_L + r_x)}{R_c}\right]$

Fig. 16. A—Series saturable reactor. B—Block diagram for series saturable reactor

eral analysis resulting in a general transfer function precludes the evaluation of these form factors. Therefore power will be defined here in terms of average volt-amperes. Thus the output power becomes $\bar{e}_L^2/R_L$, the power drawn from the signal source becomes $\bar{e}_c\bar{i}_c$, etc.

In the author's opinion the term "power gain of a magnetic amplifier" should not be so construed as to include the power dissipated in the signal source impedance. Therefore, in the following, the power gain G_p is defined as the ratio of load power to power delivered at the driving point, and the term "control circuit efficiency" η_c is defined as the ratio of power delivered at the driving point to total power delivered by the signal source. The product of the two $(G_p\eta_c)$ is therefore the ratio of load power to power delivered by the signal source.

The dynamic power gain for the circuit with both intrinsic and extrinsic negative feedback (Fig. 3) is derived in the appendix as

$$G_p = \hat{R}\left[\frac{R_L}{(r_x+R_L)^2}\right] \times \left[\frac{\epsilon^{-\tau_s}}{\left(1+\frac{r_c}{\hat{R}'}\right)-\left(1-\frac{1}{A_e}\right)\epsilon^{-\tau_s}}\right] \tag{37}$$

and the static power gain is

$$\operatorname*{Limit}_{s\to 0} G_p = \hat{R}\left[\frac{R_L}{(r_x+R_L)^2}\right]\left[\frac{1}{\frac{r_c}{\hat{R}'}+\frac{1}{A_e}}\right] \tag{38}$$

If the external feedback is adjusted for half-cycle response $(A_e=1)$,

$$G_p = \hat{R}\left[\frac{R_L}{(r_x+R_L)^2}\right]\left[\frac{1}{1+\frac{r_c}{\hat{R}'}}\right]\epsilon^{-\tau_s}, \; [A_e=1] \tag{39}$$

and the static power gain becomes

$$\operatorname*{Limit}_{s\to 0} G_p = \hat{R}\left[\frac{R_L}{(r_x+R_L)^2}\right]\left[\frac{1}{1+\frac{r_c}{R'}}\right], \quad [A_e=1] \tag{40}$$

The circuit reverts to the conventional full-wave magnetic amplifier upon the removal of the external feedback $(A_e\to\infty)$ and the dynamic power gain is

$$G_p = \hat{R}\left[\frac{R_L}{(r_x+R_L)^2}\right]\left[\frac{\epsilon^{-\tau_s}}{\left(1+\frac{r_c}{\hat{R}'}\right)-\epsilon^{-\tau_s}}\right], \quad [A_e\to\infty] \tag{41}$$

The static power gain therefore is

$$\operatorname*{Limit}_{s\to 0} G_p = \hat{R}\left[\frac{R_L}{(r_x+R_L)^2}\right]\left[\frac{\hat{R}'}{r_c}\right], \quad [A_e\to\infty] \tag{42}$$

Note that if the winding resistance r_c were zero the power gain would be infinite. Under these conditions the magnetic amplifier requires power only to change from one steady state to another.

The control circuit efficiency for the amplifier with external feedback is given by

$$\eta_c = \left[\frac{1}{1+\left[\frac{R_s}{r_c+\hat{R}'\left[1-\left(1-\frac{1}{A_e}\right)\epsilon^{-\tau_s}\right]}\right]}\right] \tag{43}$$

and

$$\operatorname*{Limit}_{s\to 0} \eta_c = \left[\frac{1}{1+\left[\frac{R_s}{r_c+\frac{\hat{R}'}{A_e}}\right]}\right] \tag{44}$$

The product $(G_p\eta_c)$ which could be termed "over-all power gain" can be determined from the foregoing to be

$$G_p\eta_c = \hat{R}\left[\frac{R_L}{(r_x+R_L)^2}\right]\left[\frac{A_iA_e}{A_i+A_e}\right]G(s) \tag{45}$$

and if $A_e=1$,

$$G_p\eta_c = \hat{R}\left[\frac{R_L}{(r_x+R_L)^2}\right]\left[\frac{1}{1+\frac{1}{A_i}}\right]G(s), \quad [A_e=1] \tag{46}$$

Likewise, the removal of the external feedback $(A_e\to\infty)$ yields

$$G_p\eta_c = \hat{R}\left[\frac{R_L}{(r_x+R_L)^2}\right][A_i]G(s), \quad [A_e\to\infty] \tag{47}$$

The figure of merit M is defined as the ratio of the static over-all power gain $(G_p\eta_c)$ to the response time (τ_a). The following is derived in the appendix.

$$M = \hat{R}\left[\frac{R_L}{(r_x+R_L)^2}\right]\left[\frac{1}{1+\frac{1}{A_i}}\right]\left[\frac{\omega_s}{\pi}\right] \tag{48}$$

Although the derivation of the foregoing expression indicates that M is independent of external voltage feedback, it can be shown that M is not independent of external current feedback. For example, since feedback into a separate winding is characteristically current feedback, the figure of merit of the multiple feedback loop circuit of Fig. 14 is modified by the external feedback. In particular, M is given by

$$M=\hat{R}\left[\frac{1}{R_L}\right]\left[\frac{1}{1+1/A_t}\right]\times \left[\frac{\omega_e}{\pi}\right]\left[\frac{1}{1-(N_b/N_x)(\hat{R}/R_b)}\right] \quad (49)$$

It is interesting to note that the previously mentioned positive feedback adjustment (i.e., $R_b=\hat{R}N_b/N_x$) results in an infinite figure of merit. Thus the self-balancing potentiometer type of circuit utilizes positive current-feedback to make M infinite and negative voltage feedback to provide stability without affecting M.

Block-Diagram Analysis Applied to Saturable Reactors

The techniques outlined in this paper may be readily applied to the parallel saturable reactor, to the series saturable reactor with a low impedance control circuit, and with some reservations to the series saturable reactor in general. Although the detailed block-diagram analysis of the saturable reactor is the subject of a later paper, the block diagrams for the parallel and for the series saturable reactor are presented in this paper for the sake of completeness. It is hoped that these diagrams, Figs. 15 and 16, will be of some value to those who are pursuing other methods of analysis of the subject circuits.

It is apparent from the relative complexity of the block diagrams that the saturable reactor does not afford the easiest starting point for the general analysis of magnetic amplifiers. The block diagrams reveal that the saturable reactor is a magnetic amplifier which is characterized by 100% intrinsic, negative current feedback. This intrinsic feedback both raises the steady-state output impedance to make the saturable reactor a current source, and lowers the input impedance of the component saturable transformers.

The various possible modes of operation in the series saturable-reactor prevent the rigorous application of the block-diagram in anything but the "low control circuit impedance" case. The investigation of the saturable reactor in its various operating modes has been undertaken by Barker, Storm, and others.[13,14]

Conclusions

A block-diagram method for treating magnetic amplifiers has been presented. It has been shown that both half-cycle response amplifiers and conventional magnetic amplifiers can be readily analyzed by this method. The time constant normally associated with magnetic amplifiers has been shown to be a result of intrinsic positive feedback which may be nullified to produce half-cycle response. In addition, the conditions under which a magnetic amplifier may be treated as a voltage-controlled device and those under which it is better treated as a current-controlled device are described. The techniques outlined in this paper provide a compatible framework within which investigations of dynamic core behavior and rectifier unblocking may be performed. The block-diagram method of analysis has been found to be particularly valuable in the study of complex magnetic amplifier circuits with multiple feedback loops.

Appendix

Development of Block Diagram of Fig. 2

The voltage loop equation for the control circuit (Fig. 1) during the $(n-1)$th half-cycle is

$$\bar{e}_g'(n-1)-\bar{i}_c(n-1)R_c-\bar{e}_c(n-1)+\bar{E}_b=\bar{e}_r'(n-1) \quad (50)$$

Since

$$[\bar{e}_r'(n-1)]=\left(\frac{N_c}{N_x}\right)[\bar{e}_r(n-1)]=\left(\frac{N_c}{N_x}\right)[\bar{e}_g(n)] \quad (51)$$

and

$$\left(\frac{N_c}{N_x}\right)\bar{e}_g(n)=\left(\frac{N_c}{N_x}\right)[\bar{E}_s-\bar{e}_o(n)] \quad (52)$$

the substitution of equations 51 and 52 into equation 50 yields

$$[-\bar{e}_c(n-1)+\bar{E}_b+\bar{e}_g'(n-1)-\bar{i}_c(n-1)R_c]\left[\frac{N_x}{N_c}\right]=\bar{e}_g(n)=[\bar{E}_s-\bar{e}_o(n)] \quad (53)$$

Taking the Laplace transform yields

$$[-\bar{e}_c(s)e^{-T_s}+\bar{E}_b(s)e^{-T_s}+\bar{e}_g(s)e^{-T_s}-\bar{i}_c(s)R_ce^{-T_s}]\left[\frac{N_x}{N_c}\right]=[\bar{E}_s(s)-\bar{e}_o(s)] \quad (54)$$

or

$$[-\bar{e}_c(s)+\bar{E}_b(s)+\bar{e}_g'(s)-\bar{i}_c(s)R_c]\left[\frac{N_x}{N_c}\right]e^{-T_s}=[\bar{E}_s(s)-\bar{e}_o(s)] \quad (55)$$

Since

$$\bar{i}_c(s)R_c=\bar{e}_r'(s)[R_c/\hat{R}']=\bar{e}_r'(s)\left[\frac{1}{A_t}\right]$$

the above becomes

$$[-\bar{e}_c(s)+\bar{E}_b(s)+\bar{e}_g'(s)]\left[\frac{1}{1+\frac{1}{A_t}}\right]\left[\frac{N_x}{N_c}\right]e^{-T_s}=[\bar{E}_s(s)-\bar{e}_o(s)] \quad (56)$$

Equations 55 and 56 are obviously identical to those that would be obtained directly from the block diagram of Fig. 2.

Simplification of Expression for Break Frequency ω_0

Given:

$$\frac{\omega_0}{\omega_e}=\frac{1}{2\pi}\sin^{-1}\left[\frac{2A_0^{-1}}{A_0^2}\right] \quad (57)$$

or

$$\frac{\omega_0}{\omega_e}=\frac{1}{2\pi}\sin^{-1}\left[\frac{2}{A_0}\left(1-\frac{1}{2A_0}\right)\right]$$

and $A_0\geq 10$, it follows that

$$\frac{\omega_0}{\omega_e}\simeq\frac{1}{2\pi}\sin^{-1}\left[\frac{2}{A_0}\right], \quad [A_0\geq 10]$$

Since for small angles, the angle is equal to its sine,

$$\frac{\omega_0}{\omega_e}\simeq\left[\frac{1}{\pi}\right]\left[\frac{1}{A_0}\right], \quad [A_0\geq 10] \quad (58)$$

Derivation of Driving Point Impedance

The driving point impedance for the circuit of Fig. 3 is given by

$$Z_t=\frac{\Delta\bar{e}_c}{\Delta\bar{i}_c}-R_s=\frac{\Delta\bar{e}_c}{\Delta\bar{i}_c}-(R_c-r_c)$$

and therefore

$$(Z_t-r_c)=\frac{\Delta\bar{e}_c}{\Delta\bar{i}_c}-R_c \quad (59)$$

Since

$$\Delta\bar{i}_c=\frac{\Delta\bar{e}_r'}{\hat{R}'}\text{ and }(\bar{e}_r')(N_x/N_c)e^{-T_s}=\bar{e}_o$$

then

$$(Z_t-r_c)=\frac{\hat{R}'}{(\Delta\bar{e}_o/\Delta\bar{e}_c)(N_x/N_c)}e^{-T_s}-R_c \quad (60)$$

Since from equation 3

$$\frac{\bar{e}_o}{\bar{e}_c}=\frac{N_x}{N_c}\left[\frac{e^{-T_s}}{\left(1+\frac{1}{A_t}\right)-\left(1-\frac{1}{A_e}\right)e^{-T_s}}\right].$$

it follows that

$$(Z_t-r_c)=\hat{R}'\left[1-\left(1-\frac{1}{A_e}\right)e^{-T_s}\right]$$

and therefore

$$Z_t = r_c + \hat{R}'\left[1-\left(1-\frac{1}{A_e}\right)\epsilon^{-T_s}\right] \qquad (61)$$

Power Gain Derivation

The magnetic amplifier power gain can be determined as follows. Since

$$G_p = \frac{\bar{e}_L^2/R_L}{\left[\bar{e}_c\left(\frac{Z_t}{Z_t+R_s}\right)\right]^2/Z_t} \qquad (62)$$

$$Z_t = r_c + \hat{R}'\left[1-\left(1-\frac{1}{A_e}\right)\epsilon^{-T_s}\right]$$

and

$$\bar{e}_L = \bar{e}_0\left[\frac{R_L}{R_L+r_z}\right]$$

it follows that

$$G_p = \left[\frac{\bar{e}_0}{\bar{e}_c}\right]^2\left[\frac{R_L}{(r_z+R_L)^2}\right]\frac{\left[R_s+r_c+\hat{R}'\left[1-\left(1-\frac{1}{A_e}\right)\epsilon^{-T_s}\right]\right]^2}{\left[r_c+\hat{R}'\left[1-\left(1-\frac{1}{A_e}\right)\epsilon^{-T_s}\right]\right]} \qquad (63)$$

Since

$$\frac{\bar{e}_0}{\bar{e}_c} = \frac{N_x}{N_c}\left[\frac{\epsilon^{-T_s}}{\left(1+\frac{1}{A_t}\right)-\left(1-\frac{1}{A_e}\right)\epsilon^{-T_s}}\right]$$

$$A_t = \frac{\hat{R}'}{R_c}$$

and $R_c = R_s + r_c$

it follows that

$$G_p = \hat{R}\left[\frac{R_L}{(r_z+R_L)^2}\right]\left[\frac{\epsilon^{-T_s}}{\left(1+\frac{r_c}{\hat{R}'}\right)-\left(1-\frac{1}{A_e}\right)\epsilon^{-T_s}}\right]\epsilon^{-T_s} \qquad (64)$$

Neglecting the half-cycle delay due to the mechanism of control yields equation 65:

$$G_p = \hat{R}\left[\frac{R_L}{(r_z+R_L)^2}\right]\left[\frac{\epsilon^{-T_s}}{\left(1+\frac{r_c}{\hat{R}'}\right)-\left(1-\frac{1}{A_e}\right)\epsilon^{-T_s}}\right] \qquad (65)$$

Derivation of Figure of Merit

The figure of merit M is defined as the ratio of static over-all power gain to response time.

Therefore

$$M = \frac{G_p\eta_c}{\tau_a} \qquad (66)$$

or

$$M = \frac{G_p\eta_e}{A_0 T}$$

It has been shown that

$$A_0 = \frac{A_t+1}{(A_t/A_e)+1}$$

and therefore

$$M = \frac{G_p\eta_c(A_t/A_e+1)}{T(A_t+1)} \qquad (67)$$

Since

$$G_p\eta_c = \hat{R}\left[\frac{R_L}{(r_z+R_L)^2}\right]\left[\frac{A_tA_e}{A_t+A_e}\right]$$

it follows that

$$M = \hat{R}\left[\frac{R_L}{(r_z+R_L)^2}\right]\left[\frac{1}{1+\frac{1}{A_t}}\right]\left[\frac{1}{T}\right] \qquad (68)$$

or

$$M = \hat{R}\left[\frac{R_L}{(r_z+R_L)^2}\right]\left[\frac{1}{1+\frac{1}{A_t}}\right]\left[\frac{\omega_c}{\pi}\right] \qquad (69)$$

References

1. FAST RESPONSE WITH MAGNETIC AMPLIFIERS, D. G. Scorgie. *AIEE Transactions*, vol. 72, pt. I, 1953 (Jan. 1954 section), pp. 741–49.

2. ANALYSIS OF MAGNETIC AMPLIFIERS BY THE USE OF DIFFERENCE EQUATIONS, P. R. Johannessen. *Ibid.*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 700–11.

3. DESIGN FACTORS FOR MAGNETIC AMPLIFIERS, K. W. Gardiner. *Report 7002–9*, Servomechanisms Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., Aug. 1954.

4. DYNAMIC CORE BEHAVIOR AND MAGNETIC AMPLIFIER PERFORMANCE, L. A. Finzi, D. L. Critchlow. *AIEE Transactions*, vol. 76, pt. I, May 1957, pp. 229–40.

5. THEORY OF MAGNETIC AMPLIFIERS WITH SQUARE-LOOP CORE MATERIALS, H. F. Storm. *Ibid.*, vol. 72, pt. I, Nov. 1953, pp. 629–40.

6. TRANSISTOR SWITCHES IMPROVE MAGNETIC-AMPLIFIER PERFORMANCE, R. L. Bright, R. O. Decker, R. A. Ramey. *Electrical Manufacturing*, New York, N. Y., Apr. 1955.

7. THE PULSE STRETCH COUPLING CIRCUIT, H. W. Patton. *AIEE Transactions*, vol. 75, pt. I, Sept. 1956, pp. 377–79.

8. ON THE MECHANICS OF MAGNETIC AMPLIFIER OPERATION, Robert A. Ramey. *Ibid.*, vol. 70, pt. II, 1951, pp. 1214–23.

9. MAGNETIC AMPLIFIERS OF THE SELF-BALANCING POTENTIOMETER TYPE, W. A. Geyger. *Ibid.*, vol. 71, pt. I, 1952 (Jan. 1953 section), pp. 383–95.

10. FULL-WAVE REVERSIBLE-POLARITY HALF-CYCLE-RESPONSE MAGNETIC AMPLIFIERS, C. B. House. *Ibid.*, vol. 74, pt. I, Sept. 1955, pp. 541–52.

11. AN OPERATIONAL-TYPE MAGNETIC AMPLIFIER FOR AIR-BORNE SERVO-CONTROL SYSTEMS, R. M. Hubbard. *Ibid.*, vol. 73, pt. I, Nov. 1954, pp. 425–30.

12. THE OPERATION OF THE SELF-BALANCING MAGNETIC AMPLIFIER, Albert D. Krall, E. T. Hooper. *Ibid.*, vol. 75, pt. I, Mar. 1956, pp. 79–84.

13. THE SERIES MAGNETIC AMPLIFIER: PART I—THE FOUR MODES OF OPERATION; PART II—THE MECHANISMS OF GAIN AND TIME CONSTANT, R. C. Barker. *Ibid.*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 819–31.

14. MAGNETIC AMPLIFIERS (book), H. F. Storm et al. John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 113–24.

Cycle- and Delay-Time Considerations in a Real-Time Digital Computer

HERBERT FREEMAN
ASSOCIATE MEMBER AIEE

Synopsis: The use of digital computers in physical control systems requires consideration of certain special effects not normally encountered with analog computers, namely, the sampling action and the computation delay. In this paper, the two effects are defined and their relationship to the digital computer program established. It is shown how they may be selected so as to result in program optimization. A special "information-handling efficiency" criterion is developed to provide a basis for defining program quality. A general discussion is given of methods for improving the program quality by taking into account the variations in information contents among the input-output data paths in a program.

A DIGITAL computer is said to operate in "real time" when it is an inherent part of a physical system. When used in this manner, a computer receives data continuously concerning certain changing physical quantities and computes data to be used in another portion of the system. For real-time operation to succeed, the speed at which successive solutions are calculated by the computer must be fast compared to that at which the physical quantities change.

The reasons which may make it desir-

Paper 57-677, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted October 22, 1956; made available for printing June 11, 1957.

HERBERT FREEMAN is with the Sperry Gyroscope Company, Great Neck, N. Y.

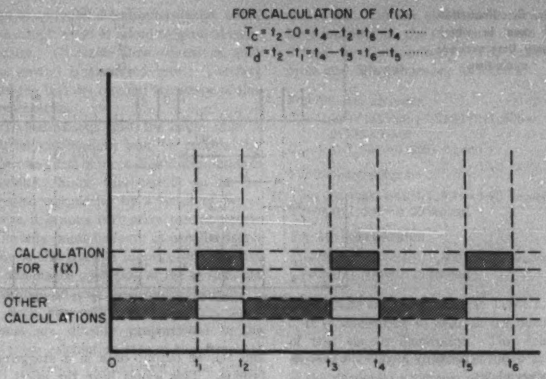

FOR CALCULATION OF f(X)
$T_C = t_2 - 0 = t_4 - t_2 = t_6 - t_4 \cdots\cdots$
$T_d = t_2 - t_1 = t_4 - t_3 = t_6 - t_5 \cdots\cdots$
CALCULATION FOR f(X)
OTHER CALCULATIONS
0
t_1
t_2
t_3
t_4
t_5
t_6

Fig. 1. Illustration of difference between cycle and delay times

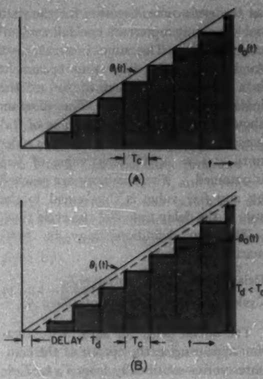


Fig. 2. Illustration of effect of delay time

able to incorporate a digital computer in a physical system are:

1. The desired output data may be such a complex function of the input data that there is no other practical method for achieving the desired result.
2. The precision required in the output data may not be realizable by any other techniques.
3. The digital computer may be simpler, smaller, lighter, less expensive, and require less power to operate than its analog counterpart. This is, of course, likely to be true only for the more complicated systems.

A digital computer is a numerical type of computing machine. If a digital computer is to be part of a physical system, it must be able to generate "numbers" from measured physical quantities. It can then carry out the sequence of arithmetical operations on these numbers as indicated by the stored program and arrive at a numerical solution. The solution must in turn be converted from a number into a physical quantity containing the same information in analog form. A digital computer when used in real time must be accompanied by an input analog-to-digital conversion device and an output digital-to-analog conversion device.

The operation of a digital computer in a physical system is roughly as follows: At a certain instant of time, the analog-to-digital converters (samplers) take readings of the instantaneous input quantities. These are converted into equivalent numerical quantities and delivered to the computer. The computer commences to solve for the answers by starting with the first instruction of the stored program and then continues in sequence until all the instructions are executed and the answers have been calculated. These are then transferred to the digital-to-analog converters and "hold" circuits where physical quantities containing the same information are generated and held. The holding operation is most important because the computer delivers a set of answers only once per cycle, whereas the output must be available continuously. As soon as the computer has calculated a set of answers, the entire cycle is repeated. It is clear that if the input quantities change excessively during the time of one cycle, a great information loss will occur from input to output and the system may be unworkable. In the other extreme, if the changes of the data during one cycle time are negligible, the digital computer will appear as an essentially continuous device, and together with its input and output devices will be indistinguishable from its analog counterpart.

The permissible cycle time is entirely a function of the rate at which the information quantities of the system change. Thus a computer working out a navigation problem aboard an ocean liner could well take an hour for a computation cycle. In contrast, a computer controlling an antiaircraft gun might have to run through its complete problem in a fraction of a second.

Definition of Cycle and Delay Times

A complicated physical system, which is the kind where a digital computer is most likely to be found, will normally have many independent variables which enter into the computations. Conceivably there may also be a number of separate outputs. If one considers a single output quantity, there are two separate time intervals that are of importance. The first consists of the time between successive instants at which new output data become available, and the second refers to the "newness" of output information when it is delivered to the system. It is desirable to make the following specific definitions:

Cycle Time. The term cycle time is used to denote the time interval between two successive calculations of the same output variable.

Delay Time. The term delay time is used to describe the time interval between the instant the first of a set of input variables is obtained and the instant the corresponding output data are delivered to the system. Delay time is a measure of the "age" of the output at the instant it becomes available.

The distinction between the two terms becomes somewhat clearer by reference to Fig. 1. The case illustrated is one in which the computer periodically calculates a set of output variables. One of these, for example, may be the calculation of $f(x)$, given x. In the diagram, the computer is shown carrying out various other computations from $t=0$ to $t=t_1$. At $t=t_1$, the computer obtains the latest value of the input variable x and commences to calculate $f(x)$. The latter is obtained at $t=t_2$, whereupon the computer repeats the entire process.

The next value of $f(x)$ is obtained at $t=t_4$. The time interval t_4-t_2 is known as the cycle time, because, for the data considered, it represents one full cycle of computation. The time interval t_4-t_3 (and t_2-t_1) is the delay time because it is a measure of the age of $f(x)$ at the instant it is computed. The diagram shows very clearly that the value of $f(x)$ computed at $t = t_4$ continues to get older until at $t = t_6$ the next value of $f(x)$ is obtained. The maximum age reached by any $f(x)$ value is thus equal to the sum of the delay time and the cycle time.

The two definitions may be summarized as follows:

Delay time = input to output = T_d

Cycle time = output to output = T_c

The actual delay (or age) of information appearing at the output of the computer varies continually from T_d to T_d+T_c. The "average age" of output information is

$$T_d+T_c/2 \tag{1}$$

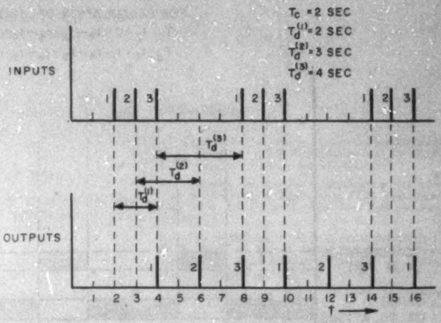
The delay time enters the problem as a pure time delay. It appears in the form ϵ^{-Ts} in those cases where the program can be described by a transfer function. The delay is caused entirely by the finite time required for the calculations of the particular input-output relation being considered and is quite similar, for example, to that encountered in an electromagnetic-delay line. A computer may have as many different delay times as it has different outputs. The delay time is directly proportional to the complexity of the particular input-output relation and inversely proportional to the speed of the computer.

The cycle time is the result of the sampling action of the digital computer and is an inherent characteristic of any sampling system. The effects of cycle time and delay time are illustrated in Fig. 2 for the case of a computer having an input-output relation of unity. A staircase-type output results from the sampling (and holding) action. If there is no delay time, the leading corners of the staircase coincide with the input. If there is a delay, the output curve is displaced to the right. For the case shown, $T_d<T_c$. This is not necessarily always the case as is shown later.

Table I. Results of Input-Output Relations

Values	Program		
	100-Step	300-Step	240-Step
Cycle time	0.153	0.92	0.46
Delay time	0.067	0.20	0.16
Maximum age	0.220	1.12	0.62
Average age	0.144	0.66	0.30
Cycle-time ratios	1	6	3
Delay-time ratios	1	3	2.4

Fig. 3. Illustration of case in which delay time exceeds cycle time



The length of the cycle time is determined by the total calculations the computer must perform within the system. Where in an analog computer the adding of a function results in an increase in equipment, in a digital computer it generally results in a longer cycle time. The cycle times need not be the same for all input-output relations; though, in general, they occur only as integral multiples of the smallest cycle time. The length of the cycle time is directly proportional to the general complexity of the entire system and inversely proportional to the speed of the computer.

As already mentioned, the delay time represents a pure time delay and is also frequently encountered in continuous systems. It does not result in a loss of information. In contrast, the cycle time represents an information loss. The greater the cycle time, the looser will be the coupling between input and output and the greater the information loss.

Normally, the delay time is less than the cycle time. This means that in general a computer solves a number of distinct problems between successive outputs of the same variable. From the definitions, the cycle time is the time from one output to the next, while the delay time is merely the time from an input to the corresponding output. The only way in which the delay time can exceed the cycle time is for an input to be received prior to the delivery of the output corresponding to the previous input. This may occur in two cases.

The first case consists of the situation in which two or more digital computers are used in parallel within a given system. (The need for this could conceivably arise from a very severe speed requirement in a highly complex system.) The computers can then accept input samples and deliver output data in a staggered fashion. This would have the effect of reducing the cycle time but leaving the delay time unchanged. For example, if a problem requires 2 seconds of computation, the use of two computers would permit accepting new input data every second and delivering output data every second. In this case the cycle time (between successive outputs) would be 1 second, and the delay time (from input to corresponding output) 2 seconds.

In the second case the inputs to the computer are already in sampled form but are delivered to the computer in an irregular manner. For example, the input data to the computer might arrive as three quantities given in fairly rapid succession, (e.g., one every second), followed by a lengthy period, e.g., 3 seconds, during which no inputs at all are received. The computer might allow 2 seconds for each calculation and deliver an output every 2 seconds. The delay times for the three input samples would then be 2, 3, and 4 seconds, respectively. See Fig. 3.

Control of Cycle and Delay Times

In the design of a digital computer for a physical control system a very important consideration is the speed of the computer. Requirements of precision and speed of response (and stability, if the computer is within a control loop) make it desirable to have a computer which is fast enough to minimize the deleterious effects of cycle time and control time. In general, the faster a computer, the greater its complexity (and cost) and, possibly, the less its reliability. Hence it is undesirable to make a computer faster than necessary. Every effort should be made to minimize the calculations required in the computer, to devise an optimum program in the computer for carrying out all the required calculations, and to arrive at a computer design which has a high ratio of speed to general complexity. Of these three objectives, only the second is discussed here. The first and the last are beyond the scope of this presentation.

It was shown that the cycle time of digital calculations can be reduced by operating two or more digital computers in parallel. Since this results in an increased complexity by a factor of two or more, it is not a very attractive approach. The only point in favor of parallel operation is the possible increase in reliability if acceptable operation of the system can continue in case one computer fails.

Both the cycle times and the delay times are directly proportional to the average calculation rate of the computer and inversely proportional to the lengths of the respective programs. In the following derivations it is assumed that the calculations are performed by a single digital computer and that the delay times are all less than their corresponding cycle times Let

D_i = the delay time corresponding to the ith input-output relation
C_i = the cycle time corresponding to the ith output
r = the average calculation rate, in program steps per second
S_i = the total number of program steps to be executed for one complete calculation of the ith input-output relation
γ_{ij} = the number of times the program steps of S_j are traversed for one traversal of S_i (note: $\gamma_{ij}=1/\gamma_{ji}$)
n = the number of input-output relations

The following can now be established

$$C_i \geqq D_i \tag{2}$$

$$D_i = \frac{S_i}{r} \tag{3}$$

$$C_i = \frac{1}{r}\sum_{j=1}^{n} \gamma_{ij} S_j (1-\delta_{ij}) + S_i \tag{4}$$

where

$$\delta_{ij} = 0 \text{ for } i \neq j \tag{5}$$

$$\delta_{ij} = 1 \text{ for } i = j$$

or

$$C_i = \frac{1}{r}\sum_{j=1}^{n} \gamma_{ij} S_j \tag{6}$$

where

$$\gamma_{ij} = 1 \text{ for } i = j \tag{7}$$

The use of equations 3 and 4 can be illustrated with the following example. Assume that a computer has an average calculation rate of 1,500 program steps per second. There are three separate input-output relations, one of 100 steps, one of 300, and one of 240, necessary to have outputs in the frequency ratio of 6 to 1 to 2, respectively. The various cycle and delay times are as follows:

For 100-step program

$T_c = [100+(1/6)300+1/3(240)]/1{,}500 = 0.153$ second
$T_d = 100/1{,}500 = 0.0667$ second

For 300-step program

$T_c = (600+300+480)/1{,}500 = 0.92$ second
$T_d = 300/1{,}500 = 0.20$ second

For 240-step program

$T_c = (300+150+240)/1{,}500 = 0.46$ second
$T_d = 240/1{,}500 = 0.16$ second

The results are summarized in Table I.

It is interesting to analyze the results of this simple illustration. The delay times are strictly a function of the lengths of the particular programs and the delay-time ratios show this very clearly. The cycle times are under the control of the programmer who here selected a 6-to-1-to 2 output frequency ratio, which of course is the inverse of the 1-to-6-to-3 cycle-time ratio. It is interesting to note that if all three programs were computed with the same cycle frequency, the single cycle time would be equal to the sum of the three delay times, i.e., 0.427 second. The example shows how it is possible to compute different programs with varying frequency, reducing some cycle times at the expense of increasing others.

Normally the problem facing the programmer will be the inverse of the foregoing. Instead of having to find the time intervals from the calculation rate and the relative cycle frequencies, the maximum permissible cycle times for the various programs will be given and the programmer will have to determine the values of γ_{ij} and the calculation rate r. The emphasis should be placed on obtaining a set of γ_{ij} values which result in a minimum value of r. On the basis of equation 6 the minimum r is obtained when the various cycle times are set as close as possible to their limiting values.

In general it is not practical to use all cycle times at exactly their maximum permissible values for the following reason. To permit an efficient split-up of programs, it is desirable that the various cycle times have some simple integral relationship to each other. A certain number of additional calculations are required when a program is interrupted and another one started. If a program is split into too many parts, these auxiliary calculations may become excessive. One can define a "major program cycle" (M) as that number of program steps with which the over-all program is periodic. The major program cycle is given simply by the sum of the program steps required to compute each of the outputs indicated by the output frequency ratio. For an efficient program split-up, the number of distinct subprograms making up the major cycle should be as small as possible.

For the example given, the relative output frequency of the three programs was 6 to 1 to 2. Let the three programs be denoted by the letters A, B, and C respectively. Program A was computed six times for one computation of B and two of C. The major program cycle is then given by:

$$\begin{aligned} M &= 6A+B+2C \\ &= 6(100)+(300)+2(240) \\ &= 1{,}380 \text{ program steps} \end{aligned}$$

Within this major program cycle, there are many ways in which the various outputs might be computed. One possible sequence might be: *c-a-a-a-c-a-b-a-a* where the *a*, *b*, and *c* are the outputs for the programs A, B, and C, respectively. The only condition imposed on the arrangement of the subprograms in the major program cycle is that the computations of the outputs of A, B, and C be precisely periodic, with the period equal to their respective cycle time.

The determination of the number of subprograms into which the basic programs must be split represents an interesting problem. Certainly each computation of an output requires a subprogram. Thus if there are six outputs for A, one for B, and two for C, there must be at least nine subprograms on this basis alone. Further, if for one of the programs A, B, or C, two outputs are computed in succession, then their computation must be separated by additional subprograms. It can thus be shown that the number of subprograms into which the main programs must be split can in no case be less than the sum of the number of outputs computed in one major cycle and the number of pairs of identical adjacent outputs. (Some of the subprograms, of course, may be identical.) In practice this theoretical minimum can usually not be achieved because of factors which depend on the actual program lengths of A, B, and C. For the example given here, the practical minimum is 13. Actual selection of the subprograms is a problem in scheduling and can be handled by various techniques associated with linear programming.

One possible arrangement for the program split-up is shown in the table on the following page. The first row identifies the subprogram, the second row gives the number of program steps it contains, and third row indicates the output which is calculated.

B_1	C_1	A_1	B_2	A_2	C_2	A_3	C_3	A_4	B_3	A_5	C_4	A_6
40	90	100	130	100	130	120	130	80	130	100	130	100
	c	a		a		a	c	a	b	a		a

There are 13 subprograms. Of these, A_1, A_2, A_5, and A_6 are equal.

Each subprogram represents a continuous program sequence capable of performing a part of the calculations for the particular input-out relation to which it refers. Within some of the subprograms, outputs may be calculated, as shown in the bottom row. This may occur at the end of a particular subprogram, or anywhere within, as the case may be. No simple relationship exists among the subprograms of a particular program; however, their sum is fixed. Thus, e.g.,

$$C_1+C_2+C_3+C_4=2C$$

Delay times were not considered for the foregoing example, except that they were not allowed to exceed their corresponding cycle times. The problem becomes considerably more complicated if the delay times are restricted to values appreciably less than the cycle times.

Program Efficiency

It has been shown how a computer can be programmed to carry out the computation of some quantities at a much higher rate than the computation of others. This fact can be used to advantage in selecting cycle times so as to optimize the information output rate of the computer. In the previous discussions it was assumed that the relative information rates of the various input-output relations were known and constant. In an actual situation this is not necessarily so, and a more general approach may have to be taken to improve what one may call the information-handling efficiency of the computer.

The information-handling efficiency of a digital computer is defined as the ratio of the total information output per unit time to the product of computer speed and word length.*

$$IH=\frac{\sum_{i=1}^{n}\frac{I_i}{C_i}}{rN} \tag{8}$$

where

IH = information-handling efficiency
I_i = information contained in the ith output, bits
C_i = cycle time of the ith output, seconds
n = total number of input-output relations
r = speed of computer, operations per second
N = word length, bits

The information-handling efficiency may be used as a measure of the quality of a computer program. For a given computer, r and N are constant and the efficiency is determined entirely by the sum of the ratios of output information to respective cycle time, I_i/C_i.

There is in general a fairly wide variation among the various I_i/C_i ratios. This arises from the fact that different input-output relations, although having the same information content, may be widely different in the complexity of the equations required to obtain the desired output data. As a result the delay times and also cycle times may differ considerably.

Combining equations 6 and 8 gives the following

$$IH=\frac{\sum_{i=1}^{n}\left(\frac{I_i}{\frac{1}{r}\sum_{j=1}^{n}\gamma_{ij}S_j}\right)}{rN} \tag{9}$$

$$=\frac{1}{N}\sum_{j=1}^{n}\frac{I_i}{\left(\sum_{i=1}^{n}\gamma_{ij}S_j\right)} \tag{10}$$

In the special case where $n=1$, this becomes

$$IH_{n=1}=\frac{I_1}{NS_1} \tag{11}$$

Equation 11 gives some insight into the meaning of the information-handling efficiency. In the special case shown here, it is directly proportional to the information contained in the output and inversely to the product of word length and number of program steps. In the limiting case, the output information makes full use of the output number, i.e., all N bits. The efficiency expression then simply reduces to $1/S_1$ and becomes equal to 100% in the limit when a full output is delivered for every single program step, a situation never attainable in practice.

It is interesting to note that the information-handling efficiency is independent of the speed of the computer, as indeed it should be since it is only a measure of the quality of a program. The concept of the information-handling efficiency is a useful one in comparing the relative merits of different computer programs and in pointing out the directions which should be followed to improve an existing program. Care must be taken to compare only programs set up for the same computer since among different computers the value of a program step may be different.

The term I_i in equation 10 represents the information content in the ith output at a given instant of time. It is information in the mathematical sense and represents the change in uncertainty due to the receipt of a particular output quantity. The information content of the output will be at a maximum of N bits if there is complete uncertainty with respect to the output quantity, i.e., if it may be any number within the capacity of the computer (N digits).

Although the concepts of delay time, cycle time, and information-handling efficiency have been developed with emphasis on the conventional arithmetic type of digital computer, they can be readily applied to the digital differential analyzer as well. Both the arithmetic type of digital computer and the digital differential analyzer handle data in quantized form and do this at discrete intervals of time, i.e., they are discrete both with respect to data and time. While the arithmetic computer determines the full values of the outputs, the digital differential analyzer merely computes increments in the output variable, one bit at a time. Hence N in equation 10 is unity for the digital differential analyzer. Since I_i may be equal to one bit, it is apparent that a digital differential analyzer is capable of a very high information-handling efficiency. The high efficiency is achieved primarily through elimination of redundancy in the output.

It is of interest to consider the case where a computer is employed to handle input-output relations having information contents which vary with time. In such a case it may be desirable to leave the selection of the optimum cycle-time ratios to the computer itself. The computer can then assign to each input-output data path a cycle time which is inversely proportional to the information currently prevailing in it. Thus an input-output data path temporarily subjected to a large information content would have a small cycle time, while another, which at the same instant contained a relatively small information content, would have a correspondingly long cycle time.

In certain control situations it may be satisfactory to have the outputs respond to the inputs merely in a general sense rather than with full precision when sudden changes in a normally slowly varying program are encountered. This fact can be used to advantage in optimizing a program by having the computer monitor the incoming data and select one of two courses of action. Normally the outputs are computed to full precision. If sudden large changes of data are detected, the computer changes the program and substitutes a short, approximate program for it (with a correspondingly smaller delay time). Simultaneously, the entire program schedule is revised to decrease the cycle time of this particular variable. This method is particularly applicable when an error signal is involved. In such a case it is generally not too important to know exactly how far the error has deviated from its zero value. It is, however, important to know the zero accurately. A program of this type could serve this purpose quite well.

---

* Word length is a term used to denote the number of binary digits which make up a number in the computer.

In the second method the computer constantly "computes ahead" during the time of the slow data variations by saying, in effect, to itself, "If the input will do this, I will do that." This can be done for a large number of possible input variations. A table of "future" outputs are then stored and kept up-to-date from cycle to cycle. If a sudden large change in input data is encountered, the computer merely selects the appropriate precalculated answer and delivers it to the output. This selection can be a very fast process, involving a minimum delay time.

The two methods just described are closely analogous to the performance of a human operator under similar circumstances. Consider, for example, an operator at the control panel in a large electric power plant. As long as everything is "running smoothly," he may carry out lengthy calculations for load distribution to keep the system performing in an optimum manner. In case of a sudden overload on one of the generators, he will temporarily put aside all considerations of optimum distribution. Instead he will concentrate on the one machine and try to correct the overload by means of quick, rough, adjustments. An alternate possibility might be to try continuously to foresee the possible difficulties which might come up and to have "answers" ready for them should they occur. These two methods are exactly the same as those previously described for the digital computer.

Conclusions

Considerable improvement in a computer program can be obtained from a design which is based on a careful selection of cycle and delay times. The improvement manifests itself in either permitting the use of a slower computer for a given problem, or in allowing the solution of a longer problem by a given computer.

The concept of information-handling efficiency provides a basis for comparing the quality of various programs and for obtaining a fuller understanding of the intrinsic information-delivering action of a digital computer.

References

1. ANALYSIS OF CONTROL SYSTEMS INVOLVING DIGITAL COMPUTERS, W. K. Linvill, J. M. Salzer. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, no. 7, July 1953, pp. 901–06.

2. FREQUENCY ANALYSIS OF DIGITAL COMPUTERS OPERATING IN REAL TIME, J. M. Salzer. *Ibid*., vol. 42, no. 2, Feb. 1954, pp. 457–66.

3. CLOSED-LOOP CONTROL SYSTEMS CONTAINING A DIGITAL COMPUTER, T. Teichmann. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, vol. EC-4, no. 3, Sept. 1955, pp. 106–17.

4. THE SYNTHESIS OF MULTIPOLE CONTROL SYSTEMS, H. Freeman. *Technical Report T-15/B* (doctoral dissertation), Electronics Research Laboratory, Columbia University, New York, N. Y., 1956.

Temperature Rise of Solid Junctions Under Pulse Load

E. J. DIEBOLD
MEMBER AIEE

Synopsis: Junctions subjected to pulse currents have a temperature rise that depends on dimensions, materials, and cooling, as much as magnitude, duration, and form of the pulse. Short pulses heat only a part of the body containing the junction; longer pulses permit the heat to penetrate deeper. Short pulses of constant magnitude, or linearly increasing in time, or capacitor discharges, show only a small difference in temperature rise. Experiments conducted with capacitor discharges are easily interpreted for other pulse shapes. For larger pulses, successive approximations are useful: plane heat flow in the junction material, plane flow in the junction and in the main body, spherical diffusion in the body, and simultaneous heating of the body with external cooling. Each approximation has its range of validity, expressed as pulse durations. Approximations overlap between ranges and cover the full span from infinitely short pulses to continuous load. A complete example is discussed.

JUNCTIONS are found in semiconductor devices or electric contacts. They are plane discontinuities in solid, homogeneous bodies. This investigation considers junctions as a finite electric and thermal resistance occurring in an infinitely thin, plane zone of a large body. Forcing electric current through the junction creates an accumulation of heat which flows into the body. The highest temperature rise occurs in the junction. High current applied for a short time (pulse) creates a high temperature difference between junction and body. Extreme temperature rise destroys the junction. The object of this investigation is to correlate magnitude and duration of current pulses with temperature rise in the junction. The results of this investigation are helpful in the design and application of junction devices.

Practical devices are too complex to fit into a simple analysis. Instead, a successive approximation assumes that various parts of the body are either not yet subjected to temperature increase, or have already reached a temperatue differential corresponding to the steady state; this reduces the problem to a simple case of heat diffusion.

Fig. 1 shows the device, its simplified form, and the following four stages of approximation. *J* is the junction, *Z* the initial zone (low conductivity materials), and *B* the body (high conductivity).

1. Initial zone only, rest cold (hence zone assumed to be infinitely large); see Fig. 1(C).

2. Initial zone at steady-state temperature differential, heat penetrates as a plane wave into the body. (Neglect spreading out of wave and external dissipation); see Fig. 1(D).

Paper 57-744, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted October 15, 1956; made available for printing April 22, 1957.

E. J. DIEBOLD is with the International Rectifier Corporation, El Segundo, Calif.

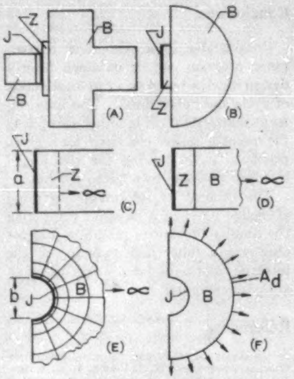


Fig. 1. Junction device and its simplified form

A—Junction device
B—Simplified form for analysis
C—Approximation by diffusion in homogeneous body
D—Plane diffusion in heterogeneous body
E—Hemispherical diffusion in homogeneous body
F—Uniform transient heating of solid mass with external heat dissipation

3. Initial and plane zone considered by assuming a reduced size junction of a hemispherical body, heat penetrates as a spherical wave into the body. (Neglect external dissipation); see Fig. 1(E).

4. Junction heating, through lumped-heat resistance, a lumped-mass body with external heat dissipation; see Fig. 1(F).

It is to be noted that an intermediary assumption between 3 and 4 would be by cylindric diffusion of heat into a cooled plate. Mathematical treatment, even of the simplest case, leads to difficulties precluding inclusion in this paper.

Although inaccurate, the successive approach shows the influence of various parts of a device, which is very helpful for the design. Furthermore, inaccuracies of mathematical approach are of lesser magnitude than the following simplifying assumptions which make the analysis possible.

1. Heat is generated only in the junction.
2. Junction infinitely thin.
3. Heat escapes only from one side of the junction.

The last assumption, although incorrect, is satisfactory if heat dissipation of the lesser part is considered in the major part by proportionally increasing the heat conductivity and the specific heat of the materials. In case of doubt, the device may be treated in stages, first for one side of the junction, then for the other side, each for an unitary value of heat dissipation. Division of heat-flow occurs in such a way that the junction temperature for both cases is the same.

Mathematical Method

Heat diffusion in a plane or spherical wave is described by two partial differential equations for temperature rise and heat-flow density as functions of distance and time (equations 1, 2, 10, 11 in Appendixes I and II). Elementary solutions, by separation of variables, have the form of equations 3, 4, or 12, 13, i.e., a product between a function of distance and a function of time, containing also an undetermined parameter (called $j\omega$) which may assume any complex value or succession of values, but independent of distance and time. These elementary solutions thus have the form of an integrand of a Fourier integral if the parameter ω is taken as a frequency. The time independent part of the elementary solutions therefore is an amplitude (spectrum) function of a Fourier integral, which may be altered by introducing any desired function of the frequency to match the given initial and borderline conditions.

Knowing the heat flow density at the junction gives the original heat-flow function for the distance corresponding to the junction (zero in the plane case, junction radius for the spherical case); applying the inverse Fourier integral gives the heat-flow spectrum function for the junction. From there the spectrum functions of both the heat-flow density and the temperature rise at any distance are determined. By applying the Fourier integral to the spectrum the original functions may be found.

Equations for plane wave are given in Appendix I, for spherical wave in Appendix II. The nomenclature gives the designations and units, and Table I gives the material constants. The initial conditions are: (1) temperature rise zero for time zero and any distance; and (2) temperature rise zero for infinite distance at any time.

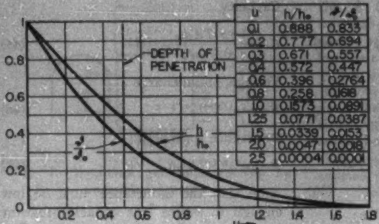
Nomenclature

h = heat flow per unit area (density), watts per square meter (w/m²)
H = amplitude function of heat density, w/m²
$H(\omega)$ = spectrum function of heat density, w/m²
H_0 = spectrum function at origin, w/m²
Q = heat flow (per unit time), watts
ϑ = temperature rise, degrees centigrade (C)
Θ = amplitude of temperature, C
$\Theta(\omega)$ = spectrum function of temperature, C
t = time, seconds
T = pulse duration, seconds
ω = frequency (parameter), 1/seconds
x = linear co-ordinate, meters (m)
x_s = thickness of initial zone, m
a = diameter of plane junction, m
b = diameter of spherical junction, m
d = depth of penetration, m
D = equivalent diameter of body, m
A_j = area of junction, m²
A_d = area of heat dissipation, m²
u = normalized variable
$\Phi(u)$ = error integral $= (2/\sqrt{\pi})\,{}_0\int^u \exp(-\nu^2)\,d\nu$
$\Gamma(n+1)$ = gamma function $= {}_0\int^\infty \alpha^n \exp(-\alpha)d\alpha$
$\Phi^1(u)$ = derivative of error integral
$h_p = \Delta e S = \rho S^{1+r}$ heat density in junction, w/m²
Δe = voltage drop $= \rho S^r$ approximation of measured curve, volts
S = current density, amperes per square meter, amp/m²
ρ = resistivity (exponential), Vm^{2r}/A^r
r = exponent
σ = rate of increase of current density, A/m²s
F = form factor of pulses
R = series resistance (capacitor), ohms
E_0 = initial voltage, volts
C = capacitance, farads
q = ratio of voltages
n = ratio of transformer (capacitor side/specimen side)
z = normalized variable
m = mass, kilograms (kg)
α = heat dissipation constant, w/m² C

Table I. Material Constants

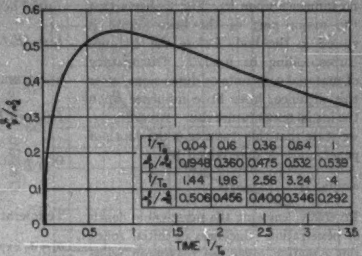
	k	c	γ	$10^{-6}\gamma c$	$10^3 p$	$10^5 \tau$
Copper	380	393	8,900	3.5	10.4	2.74
Silver	420	234	10,500	2.46	13.1	3.11
Aluminum	204	920	2,700	2.5	9.0	4.25
Nickel	58	460	8,700	4.0	3.8	6.58
Tin	64	235	7,200	1.7	5.2	9.6
Molybdenum	145	270	10,200	2.76	7.2	5
Silicon	84	735	2,340	1.72	7.0	8.3
Germanium	50	310	5,320	1.65	5.5	11
Indium	24	238	7,250	1.725	8.7	15.5
Antimony	17.5	165	6,680	1.10	4.0	22.8

k = heat conductivity, w/m C
c = specific heat, mass, ws/kg C
γ = specific mass, kg/m³
γc = specific heat, volume, ws/m³ C
$p = (k/\gamma c)^{1/2}$ propagation constant, m/s^{1/2}
$\tau = (k\gamma c)^{-1/2}$ = temperature rise constant, m²C/ws^{1/2}



u	h/h_o	ϑ/ϑ_o
0.1	0.888	0.833
0.2	0.777	0.694
0.3	0.671	0.557
0.4	0.572	0.447
0.6	0.396	0.2764
0.8	0.258	0.1618
1.0	0.1573	0.0891
1.25	0.0771	0.0387
1.5	0.0339	0.0153
2.0	0.0047	0.0018
2.5	0.0004	0.0001

Fig. 2 (left). Penetration of heat and temperature rise in a body according to Fig. 1(C). The variable u is proportional to the distance and inversely proportional to the square root of time



t/T_o	0.04	0.16	0.36	0.64	1
ϑ_p/ϑ_d	0.1948	0.360	0.475	0.532	0.539
t/T_o	1.44	1.96	2.56	3.24	4
ϑ_p/ϑ_d	0.506	0.456	0.400	0.346	0.292

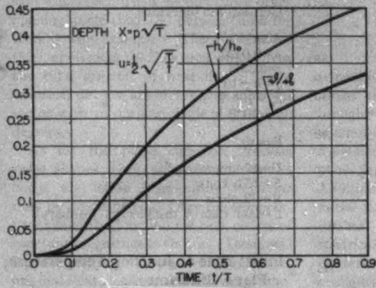


Fig. 3 (left). Increase of heat density and temperature rise at a given depth x below the junction. Time expressed in fractions of the pulse duration

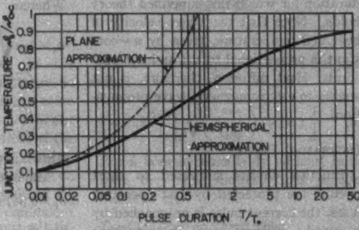


Fig. 5 (top, right). Capacitor discharge. Temperature rise of the junction against time. Equivalent pulse duration: Zero to maximum of temperature. Peak load by initial current

Fig. 6 (above). Pulse of constant magnitude in hemispherical body. Temperature rise plotted against pulse duration. Dotted line shows comparison to plane diffusion

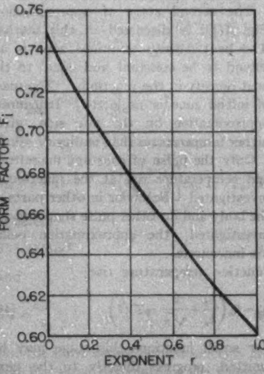


Fig. 4. Current pulse increasing linearly with time. Form factor of temperature rise as a function of the fractional exponent describing the junction voltage-current characteristic

Subscripts

p = plane junction, at the junction
s = spherical junction, at the junction
∞ = infinite time
d = capacitor discharge (heat dissipation)
i = linearly increasing pulse
c = constant pulse
1 = limit of first stage of approximation
2 = limit of second stage of approximation
3 = limit of third stage of approximation
j = junction
s = initial zone
b = body
o = origin for spectrum function, time constants

Homogeneous Body, Plane Diffusion

In this section the first stage of approximation, valid only for current pulses of very short duration, is examined. The subject is an infinitely thin junction affixed to homogeneous body having the properties of initial zone (e.g., silicon, molybdenum, tin), plane diffusion of heat. Equations pertaining to this case are shown in Appendix I.

PULSE OF CONSTANT MAGNITUDE

During the pulse, the initial heat density of the plane junction is constant. At the junction ($x=0$) equation 9 gives the spectrum function:

$$H_p = H = h_p/j\omega \tag{19}$$

According to equations 6 and 8 the junction temperature is

$$\vartheta_p = (2/\sqrt{\pi})\tau h_p t^{1/2} \tag{20}$$

Substituting the pulse duration T for the time t, this gives the highest temperature attained in any point of the body.

SPACE-TIME DISTRIBUTION OF HEAT

From equation 19, the spectrum functions 5 and 6 are determined. A table of Fourier integrals[1] gives the heat and temperature distribution. With

$$u = x/2p\sqrt{t} \tag{21}$$

$$h/h_p = 1-\Phi(u) \tag{22}$$

$$\vartheta/\vartheta_o = \exp(-u^2) - u\sqrt{\pi}[1-\Phi(u)] \tag{23}$$

Wherein ϑ_o is given by equation 20. Numerically, 23 may be rewritten:

$$2\vartheta/\sqrt{\pi}\vartheta_o = \Phi'(u) - 4u[1-\Phi(u)] \tag{24}$$

This is easier to calculate; values are shown in reference 2. Fig. 2 shows these functions, valid for any value of distance and time within the pulse duration. For constant time, e.g., end of pulse T, Fig. 2 shows directly the spatial distribution of heat and temperature. At a given distance, e.g., $x=p\sqrt{T}$, Fig. 3 gives the rise of heat and temperature with time (determined by means of Fig. 2, varying the time in equation 21).

LIMITS OF FIRST APPROXIMATION

The maximum temperature of the junction, reached at the end of the pulse (time T) is determined by heat and temperature distribution in the body. Parts which were not reached by the heat wave have no influence upon it. Fig. 2 shows that the major part of the heat (52%) is stored in the body before the distance corresponding to u=1/2. Discrepancy of material constants or dimensions, after this distance, have little influence upon the initial temperature rise. Accordingly, we state that this distance is the depth of penetration d, determined by

$$d=p\sqrt{T_1} \tag{25}$$

Inversely, knowing the permissible depth of penetration (thickness of initial zone) this equation gives the maximum pulse duration for which this simplified theory holds. Note that the phase velocity of heat and temperature, given by u=constant is proportional to x/t and not constant. Heat diffusion, therefore, is not analog to electromagnetic wave propagation!

Increasing Pulse

Junction devices subjected to short circuits are protected by fast operating fuses, the current pulse is approached by a steady increase and instantaneous cutoff. Hence current density and initial heat are

$$S=\sigma t \qquad h_p=\rho\sigma^{1+r}t^{1+r} \tag{26}$$

This gives the spectrum functions at the origin:

$$H(\omega)=\frac{\rho\sigma^{1+r}\Gamma(2+r)}{(j\omega)^{2+r}} \tag{27}$$

$$\Theta(\omega)=\frac{\tau\rho\sigma^{1+r}\Gamma(2+r)}{(j\omega)^{5/2+r}} \tag{28}$$

And the temperature rise

$$\vartheta_o=\tau\rho\sigma^{1+r}\,\frac{\Gamma(2+r)}{\Gamma\left(\frac{5}{2}+r\right)}\,t^{3/2+r} \tag{29}$$

For better comparison with equation 21 put

$$h_p=\rho\sigma^{1+r}\,T^{1+r} \tag{30}$$

$$\vartheta_p=F_t\tau h_pT^{1/2} \tag{31}$$

$$F_t=\Gamma(2+r)/\Gamma(5/2+r) \tag{32}$$

The form factor F_t is plotted in Fig. 4, which shows that the exponent r has very little influence upon the form factor (r is only an approximation of an experimental curve).

Capacitor Discharge

Experiments with short pulses use capacitor discharges. Capacitance C, charged at the voltage E_o discharged through resistance R gives the current density

$$S=(E_o/A_jR)\exp(-t/RC) \tag{33}$$

Initial heat density:

$$h_o=\rho(E_o/A_jR)^{1+r}\exp\left(-\frac{1+r}{RC}t\right) \tag{34}$$

Spectrum functions at origin

$$H(\omega)=\rho(E_o/A_jR)^{1+r}\bigg/\left(j\omega+\frac{1+r}{RC}\right) \tag{35}$$

$$\Theta(\omega)=\tau\rho(E_o/A_jR)^{1+r}\bigg/\left(j\omega+\frac{1+r}{RC}\right)(j\omega)^{1/2} \tag{36}$$

Temperature rise of the junction:

$$\vartheta_o=\vartheta_d\exp(-z^2)\int_0^z\exp y^2\,dy \tag{37}$$

Wherein:

$$\vartheta_d=2\tau\rho\sqrt{\frac{RC}{\pi(1+r)}}\left(\frac{E_o}{A_jR}\right)^{1+r} \tag{38}$$

$$z=\sqrt{t/T_o} \tag{39}$$

$$T_o=\frac{RC}{1+r} \tag{40}$$

This function is shown in Fig. 5.

Maximum temperature

$$\vartheta_p=0.541\,\vartheta_d$$

at the time

$$T=0.85\,T_o \tag{41}$$

which is (arbitrarily) called the pulse duration.

Introducing the maximum heat density (time zero)

$$h_p=\rho\left(\frac{E_0}{A_jR}\right)^{1+r} \tag{42}$$

The maximum temperature is

$$\vartheta_p=0.662\,\tau h_pT^{1/2} \tag{43}$$

Correlation of Temperature Rise for Short Pulses

Short pulses cause a temperature rise of the form:

$$\vartheta_p=(Fh_p)_\tau T^{1/2} \tag{44}$$

Wherein

F is a form factor
Constant pulses $F_c=2/\sqrt{\pi}=1.128$
Increasing pulses F_t given by Fig. 4, average 0.680
Capacitor discharge $F_d=0.662$

Suggestions for Experiments

Direct capacitor discharges require extremely high capacity at low voltage; it is more advantageous to use a pulse transformer. The discharge circuit should contain enough resistance to assure the pulse shape of a true capacitor-resistor discharge. Introducing the ratio q of initial voltage to specimen voltage (substantially larger than unity) and the ratio n of transformation of the pulse transformer ($n>1$), we obtain:

Capacity

$$C=\frac{(1+r)TA_jh_p{}^{1+r/1-r}}{0.85qn^2\rho^{2/1+r}} \tag{45}$$

Voltage, initial charge

$$E=qn\rho^{1/1+r}h_p{}^{r/1+r} \tag{46}$$

Resistance (primary)

$$R=qn^2\rho^{2/1+r}/A_jh_p{}^{1-r/1+r} \tag{47}$$

Note that the inductance must be low enough to assure true capacitor-resistor discharge.

An example follows:
Germanium junctions,

Area 10^{-4} m^2
$\rho=0.0127$
$r=0.263$
$q=n=10$
$h_p=10^7$ W/m^2
$T=10^{-3}$ s

Result:

$C=47$ microfarads
$E=561$ volts
$R=32$ ohms
$L=$less than 5 millihenrys (primary)

Initial Zone at Junction Temperature, Plane Diffusion

The second stage of approximation, Fig. 1(D) is discussed in this section. The heat density of initial zone is assumed to be constant and equal to the heat density in the junction. The mass of initial zone is neglected. Inaccurate approximation on the safe side, gives higher temperatures than reality.

Only the pulse of constant magnitude and temperature rise at the junction is investigated. Behavior in other parts of the body and for other pulse shapes is not investigated; the approximation being too inaccurate.

Junction temperature rise:

$$\vartheta_p=h_p\left(\frac{x_2}{k_2}+\frac{2}{\sqrt{\pi}}\tau_2T^{1/2}\right) \tag{48}$$

For small junctions this stage may be omitted, progress directly to the next stage of approximation by introducing the initial temperature drop as in equation 48.

Approximation by Spherical Diffusion

A hemispherical junction is affixed to a semi-infinite body, according to Fig. 1(E). Equation for heat-flow density and temperature rise are given in Appendix II. It will be noted that in this equation the amount of heat does not appear, it can be obtained by multiplying heat density and flow area. Consequently, the equations are also correct for spherical diffu-

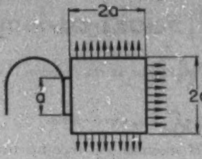
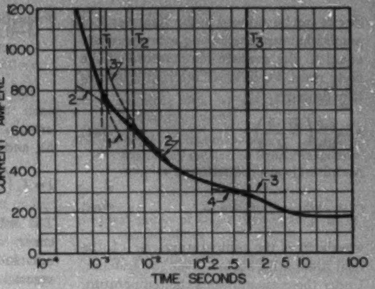


Fig. 7. Simplified germanium rectifier for numerical example

Fig. 8 (right). Current in pulses of constant magnitude, plotted against time, needed to cause a temperature rise of 40 C in the junction of Fig. 7. The dotted lines 1 to 4 indicate the stages of approximation, the times T the theoretical limits of each stage of approximation. Solid line: assumed curve for the complete device



sion which occurs in a sector of any other opening angle than 180 degrees (used for the following example). The only change, then, appears in the flow area which gives the total amount of heat flow. Many practical devices are closer to a spherical sector with an opening angle of 90 to 120 degrees.

As for the plane diffusion, various pulse shapes, their penetration into the body, and their temperature behavior may be investigated. Only the temperature rise of the junction for heat pulses of constant magnitude is shown. Designating as b, diameter of spherical junction, h_s original heat density, the following functions are obtained from Appendix II.

$$H_0=\frac{4H}{b^2}\left(1+\frac{b}{2p}\sqrt{j\omega}\right)\exp\left(\frac{-b}{2p}\sqrt{j\omega}\right)=\frac{h_s}{j\omega} \tag{49}$$

$$\Theta_0=\frac{h_s b}{2k_b}\Big/ j\omega\left(1+\frac{b}{2p}\sqrt{j\omega}\right) \tag{50}$$

Pulse duration T, temperature at origin:

$$\vartheta_s/\vartheta_\infty=1-\left[1-\Phi\left(\sqrt{\frac{T}{T_0}}\right)\right]\exp\left(T/T_0\right) \tag{51}$$

Wherein:

$$T_0=b^2/4p_b{}^2 \tag{52}$$

and

$$\vartheta_\infty=\frac{h_s b}{2k_b} \tag{53}$$

For the same hemispherical junction area and the same designations the temperature rise for the plane approximation is:

$$\vartheta_p/\vartheta_\infty=\frac{2}{\sqrt{\pi}}\sqrt{T/T_0} \tag{54}$$

Fig. 6 shows the temperature rise curves, solid line for spherical case and dotted line for plane case (compared to hemisphere).

Limit of Second Approximation (for Hemispherical Third Approximation)

For very short pulses, Fig. 6 shows that plane and spherical flow give the same temperature rise. For pulses in the range of $T=T_0/5$, both cases can be brought together by considering the same amount of initial heat in the junction, but a smaller area of the spherical than of the plane junction, i.e., higher heat density. Additional obstacles to heat transfer, i.e., initial zone, may also be considered by further reduction of the equivalent spherical junction. Thus the true configuration of Fig. 1(A) is replaced by the configuration of Fig. 1(E), in which the junction area is smaller than in reality. This assumption gives higher temperature values, because the heat storing capacity of the approximation is smaller than in the real device.

Assume equal temperature rise for the case of plane and spherical diffusion at a pulse duration:

$$T_2=T_0/5 \tag{55}$$

Heat flow densities of the plane (p), diameter a, and spherical (s), diameter b, for the same amount of initial heat are

$$h_s=(a^2/2b^2)h_p \tag{56}$$

Figure 6 gives the temperature rise

$$\vartheta_s=0.36\vartheta_\infty \tag{57}$$

Using equations 52 and 53 we obtain the spherical temperature per unit of heat density of the equivalent plane junction,

$$\vartheta_s/h_p=a^2/49.7p_bk_b\sqrt{T_2} \tag{58}$$

For the plane assumption we have (equation 48):

$$\vartheta_p/h_p=\frac{x_s}{k_s}+\frac{2}{\sqrt{\pi}}r_b\sqrt{T_2} \tag{59}$$

Equal temperature rise for both cases occurs at the pulse duration:

$$T_2=\left(\frac{0.443x_s}{r_bk_s}\right)^2\left[\sqrt{\left(\frac{0.3ak_s}{x_sk_b}\right)^2+1}-1\right]^2 \tag{60}$$

The pulse duration in turn determines the diameter of the equivalent spherical junction:

$$b=1.98\frac{x_sk_b}{k_s}\left[\sqrt{\left(\frac{0.3ak_s}{x_sk_b}\right)^2+1}-1\right] \tag{61}$$

For a homogeneous body without initial zone the pulse duration and junction diameter are

$T_2=(a/7.5p_b)^2$ and $b=0.596a$

or a depth of penetration of the plane wave $d=a/7.5$

Limit of Third Approximation (Hemisphere)

When the total heat developed in the junction causes a general temperature rise of the spherical body within its natural confinements, the third approximation ceases to be valid. Total heat developed equals heat stored in mass of body; average temperature rise assumed at 1/4 of junction temperature rise. This leads to

$$\pi a^2h_pT_3=mc\vartheta_s \tag{62}$$

$$\frac{T_3/T_0}{\vartheta_s/\vartheta_\infty}=\frac{mcp_b{}^2}{\pi k_bb^3} \tag{63}$$

Equation 63 and Fig. 6 permit us to determine the limit pulse duration (by trial and error).

Body at Equal Temperature, Heat Dissipated

Pulses of long duration and low magnitude cause slow heating of the junction, the heat has time enough to spread through the entire device under investigation. Accordingly, the fourth stage of approximation assumes that a mass m with the specific heat c is heated uniformly by the heat developed in the junction. Cooling is provided by a heat dissipating surface area A_d with the heat dissipation constant α. Assuming, further, that the outside dimensions of the device can be reduced to the diameter D of a solid hemisphere having a temperature equal to the average temperature of the body mass, we have an internal temperature drop of:

$$\vartheta_t=\frac{h_s b}{2k_b}\left(\frac{D-b}{D}\right) \tag{64}$$

An initial heat production of

$$Q_t=\frac{\pi}{2}b^2h_s=\vartheta_0\alpha A_d+mc\vartheta_0' \tag{65}$$

wherin ϑ_0 is the temperature rise of the heat dissipating surface over the ambient. Under these assumptions the temperature rise of the junction over ambient is

$$\vartheta\infty=\frac{a^2h_p}{4b}\left\{\frac{D-b}{k_bD}+\frac{\pi b}{\alpha A_d}\times\left[1-\exp\left(-\frac{T\alpha A_d}{mc}\right)\right]\right\} \tag{66}$$

For infinitely long pulses, the stationary temperature rise is

$$\vartheta\infty=\frac{a^2h_p}{4b}\left(\frac{D-b}{k_bD}+\frac{\pi b}{\alpha A_d}\right) \tag{67}$$

In the example that follows, assume a germanium diode according to Fig. 7. Heat dissipation through the anode is neglected. Heat dissipation on the surface of the cathode body may occur by forced convection with fins or by liquid cooling. The mass of the fins or cooling fluid are neglected. Pulses of constant magnitude; for each pulse duration calculate the current causing a temperature rise of 40 C.

First approximation: Heat density from equation 20. Limit of validity, equation 25.

Second approximation: Heat density from equation 48. Limit of validity, equation 60.

Third approximation: Equivalent junction diameter, equation 60. Heat density from equations 51, 52, 53 (Fig. 6). Limit of validity, equation 63 (Fig. 6).

Fourth approximation: Heat density from equation 66.

Geometric dimensions
$a=10^{-3}$ m
$A_t=7.87\times10^{-6}$ m^2
$A_d=1.57\times10^{-3}$ m^2
$x_s=2.5\times10^{-4}$ m
$D=2\times10^{-2}$ m (approximately)

Initial zone material
$k_s=70$ w/m C
$\tau_s=10^{-4}$ m^2 C/ws$^{1/2}$
$p_s=0.006$ m/s$^{1/2}$

Body material
$k_b=380$ w/m C
$\tau_b=2.74\times10^{-6}$ m^2 C/ws$^{1/2}$
$p_b=0.01$ m/s$^{1/2}$

Properties of junction
$\rho=0.0127$ Vm^{2r}/A^r
$r=0.263$

Properties of body
$m=0.056$ kg
$c=393$ ws/kg C
$\alpha=3{,}000$ w/m^2 C

Fig. 8 shows the current required for each pulse duration. Current values are determined from the equations 20, 48, 51, and 66 with the relations:

$$I=A_j\left(\frac{h_p}{\rho}\right)^{1/1+r}=A_j\left(\frac{2b^2h_s}{a^2\rho}\right)^{1/1+r} \tag{68}$$

The solid line shows the approximate curve for the real device, whereas the numbered dotted lines represent the successive approximations. Theoretical limits of validity are indicated by limit pulse durations.

Conclusions

From the curve in Fig. 8 we conclude that it must be difficult to find a single equation describing this performance and that it is unwise to extrapolate a series of test results. Inversely, test curves with several bumps may be used to interpret thermal characteristics and to suggest changes in the design.

Improvement in pulse load behavior within a certain range of approximation (given by the time limits of equations 25, 60, 63) is only possible for design changes affecting this particular, or preceding, ranges but not for those affecting any ulterior range. Briefly, heat dissipators are ineffective until the heat reaches them.

Appendix I. Plane Wave Fig. 1 (C)

Pulse starts at time zero, junction located at distance zero.

Basic equations:

$$h=-k\frac{\partial\vartheta}{\partial x} \tag{1}$$

$$\frac{\partial h}{\partial x}=-\gamma c\frac{\partial\vartheta}{\partial t} \tag{2}$$

Elementary solutions:

$$h=H\exp\left(-\frac{x}{p}\sqrt{j\omega}+j\omega t\right) \tag{3}$$

$$\vartheta=\frac{\tau H}{\sqrt{j\omega}}\exp\left(-\frac{x}{p}\sqrt{j\omega}+j\omega t\right) \tag{4}$$

Spectrum functions:

$$H(\omega)=H\exp\left(-\frac{x}{p}\sqrt{j\omega}\right) \tag{5}$$

$$\Theta(\omega)=\frac{\tau H}{\sqrt{j\omega}}\exp\left(-\frac{x}{p}\sqrt{j\omega}\right) \tag{6}$$

Fourier integral solutions:

$$h=\frac{1}{2\pi}\int_{-\infty}^{+\infty}H(\omega)\exp(j\omega t)d\omega \tag{7}$$

$$\vartheta=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\Theta(\omega)\exp(j\omega t)d\omega \tag{8}$$

Given heat pulse at the junction ($x=0$) as $h=h_p(t)$

$$H_0=H=\int_{-\infty}^{+\infty}h_p\exp(-j\omega t)dt \tag{9}$$

Pairs of spectrum and time functions are given in reference 1.

Appendix II. Hemispherical Wave Fig. 1(E)

Pulse starts at time zero, junction located at radius $b/2$.

Basic equations:

$$h=-k\frac{\partial\vartheta}{\partial x} \tag{10}$$

$$\frac{\partial h}{\partial x}+\frac{2h}{x}=-\gamma c\frac{\partial\vartheta}{\partial t} \tag{11}$$

Elementary solutions:

$$h=\frac{H}{x^2}\left(1\mp\frac{x}{p}\sqrt{j\omega}\right)\exp\left(\pm\frac{x}{p}\sqrt{j\omega}+j\omega t\right) \tag{12}$$

$$\vartheta=\frac{H}{kx}\exp\left(\pm\frac{x}{p}\sqrt{j\omega}+j\omega t\right) \tag{13}$$

Solutions are Bessel functions of the order $\pm j/2$, expressed as exponentials; positive exponents have no physical meaning.

Spectrum functions:

$$H(\omega)=\frac{H}{x^2}\left(1+\frac{x}{p}\sqrt{j\omega}\right)\exp\left(-\frac{x}{p}\sqrt{j\omega}\right) \tag{14}$$

$$\Theta(\omega)=\frac{H}{kx}\exp\left(-\frac{x}{p}\sqrt{j\omega}\right) \tag{15}$$

Fourier integral solutions:

$$h=\frac{1}{2\pi}\int_{-\infty}^{+\infty}H(\omega)\exp(j\omega t)d\omega \tag{16}$$

$$\vartheta=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\Theta(\omega)\exp(j\omega t)d\omega \tag{17}$$

Start: Given heat pulse $h=h_s(t)$ at the junction $x=b/2$

$$H_0=\frac{4H}{b^2}\left(1+\frac{b}{2p}\sqrt{j\omega}\right)\exp\left(\frac{-b}{2p}\sqrt{j\omega}\right)=\int_{-\infty}^{+\infty}h_s\exp(j\omega t)dt \tag{18}$$

References

1. FOURIER INTEGRALS FOR PRACTICAL APPLICATIONS (book), G. A. Campbell, R. M. Foster. D. Van Nostrand Company, Inc., Princeton, N. J., 1948.

2. TABLES OF FUNCTIONS WITH FORMULAE AND CURVES (book), E. Jahnke, F. Emde. Dover Publications, New York, N. Y., 1945.

Signal-to-Noise Ratios in Strong-Carrier FM Systems

HARRY URKOWITZ
NONMEMBER AIEE

Synopsis: Signal-to-noise ratios are derived for an FM (frequency modulated) system in which the r-f (radio-frequency) carrier is strong compared with the noise. The results are derived in a simple way without making use of high-order probability density functions. General formulas are derived for determining the signal-to-noise ratio with periodic modulation or with pulse modulation, and numerical results are given for the special case of a flat symmetrical r-f passband. The effect of a mistuned frequency detector is also taken into account.

THE system to be analyzed is a receiver consisting of an r-f section, a frequency detector, and an audio or video section. The r-f section is presumed to be linear. The input is modulated in frequency and the noise is linearly added to the modulated carrier in the r-f section. The frequency detector is presumed to be an ideal frequency detector whose output is proportional to the instantaneous frequency of its input. From a practical standpoint, extreme limiting corresponds very closely to the assumption of an ideal frequency detector, so that this paper analyzes a system operating with extreme limiting in the frequency detector while handling a carrier which is strong compared with the noise. The fact that extreme limiting corresponds to ideal frequency detection is shown in a paper by Middleton.[1]

Middleton's paper gives signal-to-noise ratios for weak carriers as well as strong carriers. However, his results are derived in a very general way by using higher order probability density functions. If attention is confined to the strong-carrier case, a comparatively simple derivation of signal-to-noise ratio is possible. The general formulas derived in this paper can be applied to arbitrary types of frequency modulation, i.e., sine wave, pulse, etc., provided that the signal can be properly defined after the frequency detector, and to r-f passbands of arbitrary shape, whether symmetrical or not. The carrier need not be centered at the crossover point of the frequency discriminator. Although general formulas have been derived, specific numerical results are given only for sine wave and pulse modulation and a flat, symmetrical r-f passband.

Some improvement in signal-to-noise ratio may be obtained by restricting the video bandwidth properly, but this effect is ignored, and it is assumed that the video section passes all low-frequency components of the detector output. It is found that the signal-to-noise ratio in the video depends upon the second moment of the power spectrum of the noise at the detector input. This second moment is not very sensitive to passband shape for a given 3-decibel bandwidth; the Gaussian filter, the flat filter, and the single-tuned circuit, for example, all having approximately the same value for this second moment.

In the r-f section there is a clear distinction as to what constitutes the noise and what constitutes the signal. The signal is most easily described by specifying the carrier amplitude and the amount of modulation. The noise is described by giving the shape of its power spectrum and its mean-square value. After passing through a nonlinear device such as the frequency detector, the separate contributions due to signal and noise cannot be determined. The signal and noise must be redefined in the video, and the definition differs for different types of modulation, whether sine wave, pulse, etc.

Nomenclature

A_0 = amplitude of carrier wave
B = 3-db radian bandwidth
b_n = nth moment of noise power spectrum = $(1/2\pi)\int_0^\infty(\omega-\omega_0)^n W(\omega)d\omega$
C_n = amplitude of nth noise component
$D(t)$ = modulation in frequency
D_0 = peak value of $D(t)$ for sine wave
$E_0(t)$ = video voltage
$\overline{E_0}$ = average of $E_0(t)$ over ϕ_n, the random phase of the noise
$\overline{E_0^2}$ = mean square of $E_0(t)$ over ϕ_n
$< >$ = average over modulation period for periodic modulation
k = constant of proportionality
N_1 = video noise for periodic modulation
N_2 = video noise for pulse modulation
Ps_1 = mean-square video signal for periodic modulation
S_1 = video signal for periodic modulation
S_2 = video signal for pulse modulation
$\left(\frac{S}{N}\right)_1$ = video signal-to-noise ratio for periodic modulation
$\left(\frac{S}{N}\right)_2$ = video signal-to-noise ratio for pulse modulation
$v_s(t)$ = r-f signal voltage
$v_N(t)$ = r-f noise voltage
$W(\omega)$ = r-f noise power spectrum before the frequency detector
x = r-f carrier-to-noise power ratio
Θ = total r-f phase
θ = phase shift due to total modulation (signal plus noise)
$\dot{\theta}=\frac{d\theta}{dt}$ = total instantaneous angular-frequency departure from the carrier angular frequency
$\rho^2=\frac{b_2}{b_0}$
φ = total phase of signal wave exclusive of carrier phase
$\dot{\varphi}=\frac{d\varphi}{dt}$ = instantaneous angular-frequency departure of signal wave from the carrier angular frequency
ϕ_n = random-noise phase
ω = angular frequency, radians per second
ω_0 = angular frequency of carrier
ω_d = fixed angular-frequency error due to mistuning
ω_n = angular frequency of nth noise component
$\Delta\omega$ = peak deviation due to pulse modulation

Description of FM Signal and FM Noise

If the r-f section of the receiver has a bandwidth which is narrow compared with the center frequency, the input to the frequency detector may be described by an amplitude and a phase, both of which may be functions of time. The amplitude is assumed to be constant. The total phase is given by

$$\Theta(t)=\omega_0 t+\theta(t) \tag{1}$$

where ω_0 = r-f carrier frequency in radian units, and θ = phase shift due to frequency modulation. Now

$$\dot{\theta}=\frac{d\theta}{dt} \tag{2}$$

is the total deviation from ω_0 and may include a constant departure from ω_0. The output of the frequency detector in volts is

$$E_0(t)=k\dot{\theta} \tag{3}$$

where k is a proportionality constant with the proper dimensions. In the absence of a carrier wave, $E_0(t)$ is caused only by noise. The contributions to $\dot{\theta}$ are

$$\dot{\theta}=\text{deviation due to modulation}+\text{constant frequency shift}+\text{deviation due to noise} \tag{4}$$

Paper 57-61, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted May 10, 1956; made available for printing May 20, 1957.

HARRY URKOWITZ is with the Philco Corporation, Philadelphia, Pa.

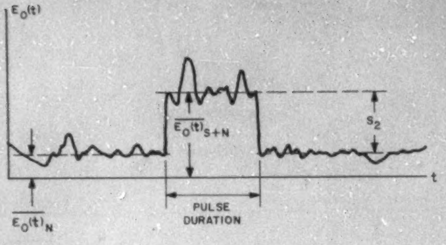


Fig. 1. Video output of frequency detector with pulse modulation

Let

ω_d = constant frequency shift
$D(t)$ = modulation as a function of time
φ = total phase difference from carrier

$$\varphi=\omega_d t+\int_0^t D(u)du \tag{5}$$

Then

$$\dot{\varphi}=\omega_d+D(t)=\text{instantaneous angular frequency} \tag{6}$$

For sine-wave modulation,

$$D(t)=D_0 \cos \omega_m t \tag{7}$$

where ω_m is the modulating frequency and D_0 represents the maximum desired deviation. The total signal voltage in the r-f section is

$$v_s(t)=A_0 \cos [\omega_0 t+\varphi(t)] \tag{8}$$

where A_0 = carrier amplitude before limiting.

For the noise in the r-f section, Rice's representation is used.[3]

$$v_N(t)=\sum_{n=1}^{\infty} C_n \cos (\omega_n t-\phi_n) \tag{9}$$

where each ϕ_n is a random phase angle uniformly distributed in the interval $(0,2\pi)$. C_n is also a random quantity and its square is of interest. C_n^2 is proportional to the mean-square voltage in the interval between the nth and the $(n+1)$ components. As the frequency interval is taken smaller and smaller the following results

$$C_n^2=\frac{W(\omega_n)\Delta\omega_n}{\pi}\rightarrow\frac{W(\omega)d\omega}{\pi} \tag{10}$$

where $W(\omega)$ is the noise-power spectrum (more precisely, the mean-square voltage spectrum), and $\Delta\omega_n$ is the radian-frequency interval between adjacent components. The mean-square noise voltage is

$$\overline{v_N^2}=\frac{1}{2}\sum_{n=1}^{\infty} C_n^2=\frac{1}{2\pi}\int_0^{\infty} W(\omega)d\omega \tag{11}$$

The total r-f voltage is now given by the equation following

$$v(t)=v_s(t)+v_N(t)$$
$$=A_0 \cos (\omega_0 t+\varphi)+\sum_{n=1}^{\infty} C_n \cos (\omega_n t-\phi_n) \tag{12}$$

Because a strong carrier is assumed, A_0 is much greater than $C_1, C_2, \ldots.$ To find the instantaneous frequency of $y(t)$, equation 38 of Appendix I can be applied to equation 12. The result is that the instantaneous angular frequency of $v(t)$ is given by

$$\dot{\Theta}=\omega_0+\dot{\phi}+\frac{1}{A_0}\sum_{n=1}^{\infty} C_n(\omega_n-\omega_0-\dot{\varphi}) \cos [(\omega_n-\omega_0)t-\varphi-\phi_n] \tag{13}$$

Then the deviation from ω_0 is

$$\dot{\theta}=\dot{\Theta}-\omega_0=\dot{\phi}+\frac{1}{A_0}\sum_{n=1}^{\infty} C_n(\omega_n-\omega_0-\dot{\varphi}) \cos [(\omega_n-\omega_0)t-\varphi-\varphi_n] \tag{14}$$

The total video output is, of course, $E_0(t)=k\dot{\theta}(t)$. The way in which the signal and the noise are defined in the video depends upon the type of modulated information that is to be extracted.

The meaning of signal and noise after the frequency detector for sine-wave modulation and for pulse modulation is now considered.

Periodic Modulation

If one is trying to extract a periodic wave, the signal is conveniently defined as $\overline{E_0(t)}$, where the bar represents an average over the random phases of the r-f noise. Then the perturbations about this phase average represent the video noise. For the periodic signal, then,

$$S_1(t)=\overline{E_0(t)} \tag{15}$$

where the subscript 1 indicates a periodic signal. For the noise

$$N_1(t)=E_0(t)-\overline{E_0(t)} \tag{16}$$

The d-c value of $S_1(t)$ is symbolized by $\langle S_1(t)\rangle$, where the symbol$\langle\rangle$ indicates a time average over the modulation period. The signal power (or rather, mean-square voltage) exclusive of direct current is

$$P_{S_1}=\langle[S_1-\langle S_1\rangle]^2\rangle$$
$$=\langle\overline{E_0(t)}^2\rangle-\langle\overline{E_0(t)}\rangle^2 \tag{17}$$

The total mean-square value of the noise voltage is obtained by averaging the square of equation 16 over the random phases of the r-f noise and over the modulation period. The result is

$$\langle\overline{N_1^2(t)}\rangle=\langle\overline{E_0^2(t)+\overline{E_0(t)}^2}\rangle$$
$$=\langle\overline{E_0^2(t)}\rangle-\langle\overline{E_0(t)}^2\rangle \tag{18}$$

The video signal-to-noise ratio is

$$\left(\frac{S}{N}\right)_1=\frac{P_{S_1}}{\langle\overline{N_1^2}\rangle}=\frac{\langle\bar{\dot{\theta}}^2\rangle-\langle\bar{\dot{\theta}}\rangle^2}{\langle\overline{\dot{\theta}^2}\rangle-\langle\overline{\dot{\theta}2}\rangle} \tag{19}$$

Pulse Modulation

A pulse-modulated FM system is now considered where a single pulse is applied. To define the S/N, an idea put forth by Jordan is extended.[3-5] The video output is illustrated in Fig. 1. In the absence of modulation due to the pulse input, the input to the discriminator consists of the noise plus the carrier whose frequency is $\omega_0+\omega_d$. This occurs, for example, during most of the sweep cycle in a radar system operating in a target-free area. The average output of the frequency detector is then $k\overline{\dot{\theta}_N}$, where the subscript N indicates that only noise is present besides the carrier. For the pulse duration, the carrier frequency becomes $\omega_0+\omega_d+\Delta\omega$, and a new value of average output appears; namely, $k\overline{\dot{\theta}_{S+N}}$. The video signal is defined as the difference between the average values with and without the pulse.

$$S_2=\overline{E_0(t)}_{S+N}-\overline{E_0(t)}_N=k(\bar{\dot{\theta}}_{S+N}-\bar{\dot{\theta}}_N) \tag{20}$$

The noise in the video is the fluctuation in the absence of a signal.

$$\overline{N_2^2}=k^2(\overline{\dot{\theta}_N^2}-\overline{\dot{\theta}_N}^2) \tag{21}$$

The signal-to-noise ratio is then defined as the ratio of the square of equation 20 to equation 21.

$$\left(\frac{S}{N}\right)_2=\frac{(\bar{\dot{\theta}}_{S+N}-\bar{\dot{\theta}}_N)^2}{\overline{{}_N\dot{\theta}^2}-\overline{\dot{\theta}_N}^2} \tag{22}$$

The signal-to-noise ratio represented by equation 22 is of a different nature than that defined by equation 19. In the case of pulse modulation, one is not interested in average power so that time averages are not taken. The criterion for pulse discernibility is the deflection from the noise average value, so that equation 22 is said to be the signal-to-noise ratio on the basis of the deflection criterion.[5]

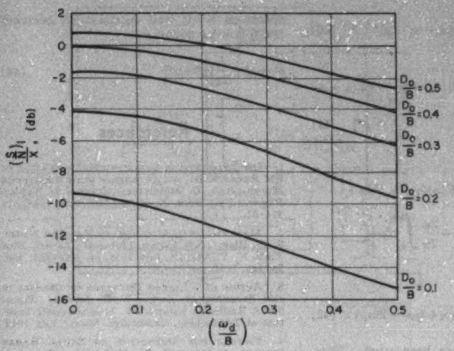


Fig. 2. Video S/N ratio in an FM system with sine-wave modulation (equation 37)

D_0 = maximum deviation, radian units
B = r-f radian bandwidth
x = r-f S/N ratio
ω_d = mistuning of frequency detector, radian units

General Video Signal-to-Noise Ratios

The values for $\bar{\dot\theta}$ and $\overline{\dot\theta^2}$ must be found to substitute into equations 19 and 22. It can be seen from equation 14 that

$$\bar{\dot\theta}=\varphi=\omega_d+D(t) \tag{23}$$

For periodic modulation $D(t)$ is a periodic function of time, while for pulse modulation $D(t)$ is the pulse function. In particular, for pulse modulation

$$\bar{\dot\theta}_N=\omega_d \tag{24}$$

and

$$\bar{\dot\theta}_{S+N}=\omega_d+\Delta\omega \tag{25}$$

where

$\Delta\omega$ = the pulse deviation

The determination of $\overline{\dot\theta^2}$ is more complicated and is carried out in detail in Appendix II. Let the noise power spectrum at the input to the frequency detector be $W(\omega)$ and the nth moment of the power spectrum be defined in the following manner[1]

$$b_n=\frac{1}{2\pi}\int_0^\infty(\omega-\omega_0)^n W(\omega)d\omega \tag{26}$$

where ω_0 is the carrier frequency. The ratio of b_2 to b_0 is designated as follows

$$\rho^2=\frac{b_2}{b_0} \tag{26A}$$

The mean-square value of the noise is, of course, b_0. Then, from Appendix II, it is found that

$$\overline{\dot\theta^2}=\varphi^2+\frac{1}{2x}\left(\rho^2+\varphi^2-\frac{2b_1\varphi}{b_0}\right) \tag{27}$$

where x = r-f carrier-to-noise power ratio. The results desired may now be obtained.

PERIODIC MODULATION

Using equations 23 and 27 to substitute into equation 19, one obtains, remembering that $D(t)$ is periodic for this case, the following equation

$$\left(\frac{S}{N}\right)_1=\frac{2x\langle D^2(t)\rangle}{\rho^2+\omega_d^2+\langle D^2(t)\rangle-2b_1\omega_d/b_0} \tag{28}$$

PULSE MODULATION

In the absence of pulse modulation

$$\varphi=\varphi_N=\omega_d \tag{29}$$

so that

$$\overline{\dot\theta_N^2}=\omega_d^2+\frac{1}{2x}(\rho^2+\omega_d^2-2b_1\omega_d/b_0) \tag{30}$$

Then, substituting equations 25, 24, and 30 into equation 22, one obtains

$$\left(\frac{S}{N}\right)_2=\frac{2x(\Delta\omega)^2}{\rho^2+\omega_d^2-2b_1\omega_d/b_0} \tag{31}$$

Special Case of Flat Symmetrical Passband

If there is a flat r-f passband extending from $\omega_0-B/2$ to $\omega_0+B/2$, then

$$\rho^2=\frac{B^2}{12} \tag{32}$$

and

$$b_1=0 \tag{33}$$

Further, for the case of periodic modulation, let the modulation be given by equation 7. Then equation 28 becomes, for sine-wave modulation,

$$\left(\frac{S}{N}\right)_1=\frac{12x(D_0/B)^2}{1+12(\omega_d/B)^2+6(D_0/B)^2} \tag{34}$$

For the passband assumed, equation 31 becomes, for pulse modulation,

$$\left(\frac{S}{N}\right)_2=\frac{24x(\Delta\omega/B)^2}{1+12(\omega_d/B)^2} \tag{35}$$

A plot of $(S/N)_1/x$, as given by equation 34, versus ω_d/B is shown in Fig. 2, and a plot of $(S/N)_2/x$, as given by equation 35, versus ω_d/B is shown in Fig. 3.

Appendix I. Frequency of Resultant of a Number of Components

Consider the sum of a number of sinusoidal components

$$R=A_0\cos\alpha_0+A_1\cos\alpha_1+A_2\cos\alpha_2+\ldots \tag{36}$$

Let A_0 be very large as compared with A_1, A_2, ..., and let α_0, α_1, ... be functions of time. By extending a formula given by Goldman,[6] the total instantaneous phase of R can be found. This is

$$\Theta=\alpha_0+\frac{A_1}{A_0}\sin(\alpha-\alpha_0)+\frac{A_2}{A_0}\times \sin(\alpha_2-\alpha_0)+\ldots \tag{37}$$

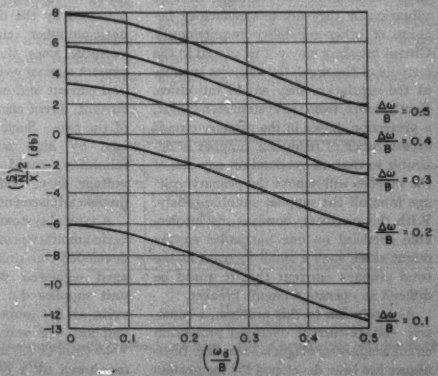


Fig. 3. Video S/N ratio in an FM system with pulse modulation (equation 38)

$\Delta\omega$ = maximum deviation, radian units
B = r-f radian bandwidth
x = r-f S/N ratio
ω_d = mistuning of frequency detector, radian units

The instantaneous angular frequency is the time derivative of equation 37 and is given by the following

$$\dot{\Theta}=\dot{\alpha}_0+\frac{A_1}{A_0}(\dot{\alpha}_1-\dot{\alpha}_0)\cos(\alpha_1-\alpha_0)+\frac{A_2}{A_0}(\dot{\alpha}_2-\dot{\alpha}_0)\cos(\alpha_2-\alpha_0)+\dots \tag{38}$$

Appendix II. Mean-Square Value of θ

From equation 14

$$\overline{\theta^2}=\varphi^2+\frac{2\varphi}{A_0}\sum_{n=1}^{\infty}C_n(\omega_n-\omega_0-\varphi)\times\overline{\cos[(\omega_n-\omega_0)t-\varphi-\theta_n]}+\frac{1}{A_0^2}\left\{\overline{\sum_{n=1}^{\infty}C_n(\omega_n-\omega_0-\varphi)\cos[(\omega_n-\omega_0)t-\varphi-\theta_n]}\right\}^2 \tag{39}$$

When averaged over ϕ_n, the result is as given in the equation

$$\overline{\theta^2}=\varphi^2+\frac{1}{2A_0^2}\sum_n C_n^2(\omega_n-\omega_0-\varphi)^2 \tag{40}$$

$$=\varphi^2+\frac{1}{2A_0^2}\sum_n C_n^2[(\omega_n-\omega_0)^2-2\varphi(\omega-\omega_0)+\varphi^2] \tag{41}$$

In the limit

$$\overline{\theta^2}=\varphi^2+\frac{1}{A_0^2}\left[\frac{1}{2\pi}\int_0^{\infty}W(\omega)(\omega-\omega_0)^2d\omega+\frac{\varphi^2}{2\pi}\int_0^{\infty}W(\omega)d\omega-\frac{2\varphi}{2\pi}\int_0^{\infty}(\omega-\omega_0)W(\omega)dW\right]$$

$$=\varphi^2+\frac{1}{A_0^2}(b_2+b_0\varphi^2-2b_1\varphi) \tag{42}$$

where equation 26 has been used. Now, let $x=A_0^2/(2b_0)$ = r-f carrier-to-noise power ratio, and $\rho^2=b_2/b_0$. Then, equation 42 becomes

$$\overline{\theta^2}=\varphi^2+\frac{1}{2x}(\rho^2+\varphi^2-2b_1\varphi/b_0) \tag{43}$$

For the special case of a symmetrical spectrum $b_1=0$, and equation 43 becomes the following equation

$$\overline{\theta^2}=\varphi^2+\frac{1}{2x}[\rho^2+\varphi^2] \tag{44}$$

References

1. ON THEORETICAL SIGNAL-TO-NOISE RATIOS IN FM RECEIVERS: A COMPARISON WITH AMPLITUDE MODULATION, **D. Middleton.** *Journal of Applied Physics*, New York, N. Y., vol. 20, Apr. 1949, pp. 334–51.

2. MATHEMATICAL ANALYSIS OF RANDOM NOISE, **S. O. Rice.** *Bell System Technical Journal*, New York, N. Y., vol. 23, July 1944, pp. 282–332; vol. 24, Jan. 1945, pp. 46–136.

3. ACTION OF A LINEAR DETECTOR ON SIGNALS IN THE PRESENCE OF NOISE, **W. H. Jordan.** *Report 61-23*, Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., July 1942.

4. FILTERS FOR DETECTION OF SMALL RADAR SIGNALS IN CLUTTER, **H. Urkowitz.** *Journal of Applied Physics*, vol. 24, pp. 1024–31, Aug. 1953, sec. 6.

5. THRESHOLD SIGNALS (book), **J. L. Lawson, G. E. Uhlenbeck.** *Radiation Laboratory Series No. 25*, McGraw-Hill Book Company, Inc., New York, N. Y., 1950, p. 161.

6. FREQUENCY ANALYSIS, MODULATION AND NOISE (book), **S. Goldman.** McGraw-Hill Book Company, Inc., 1948, p. 164.

A Magnetic Amplifier Voltage-Current Overload Protection Circuit

R. J. DISTLER
ASSOCIATE MEMBER AIEE

C. STURZENBECKER
ASSOCIATE MEMBER AIEE

LOW-VOLTAGE, high-current, regulated power supplies, such as are used in transistor computers, present an overload protection problem. In many instances fuses are prohibited because of their susceptibility to damage through vibration or they are unsatisfactory because of other reliability requirements. Circuit breakers may not be placed in the outputs of the regulated supplies because of the regulation they would introduce. Furthermore, from the system standpoint, a protection device in the primary circuit of a number of regulated supplies is desirable so that an overload on any one of the supplies will remove the primary voltage from all the supplies simultaneously. With this circuitry, however, an intolerable overload on one particular supply might cause such a small increase in the total primary currrent that it would be difficult to operate a circuit breaker.

As a solution to this problem, a sensitive voltage-current overload protection circuit employing magnetic-amplifier techniques has been developed for the protection of electric equipment. This circuit provides protection without the use of relays, fuses, or circuit breakers by using bistate magnetic amplifiers as detectors, and a self-saturating magnetic amplifier as an input voltage suppression device which is operated by the detectors. These magnetic-amplifier circuits are extremely versatile, since they can be designed to protect against overloads of both voltage and current and can monitor alternating current, direct current, or a combination of the two. Such circuits are characterized by high sensitivity, ruggedness, reliability, rapid response, low series voltage drop, and accurate and reproducible adjustment of the overload point.

A typical system for which such a magnetic-amplifier protective circuit has been developed contains four separately packaged regulated power supplies. Each unit supplies d-c loads as follows: +6 volts at 3.0 amperes, −2 volts at 2.5 amperes, −8 volts at 4.0 amperes, and −45 volts at 3.0 amperes. Each load is monitored by a current and voltage detector with the outputs of all detectors coupled to one voltage suppressor in the common 400-cps (cycles per second) power circuit. With variations in the system's 400-cps 115-volt power of ±5% for voltage and frequency, the current detectors used in this case can be preset to trip at any desired overload with an accuracy of ±0.1 ampere. For example, in the case of a −8-volt 4.0-ampere load, with the detector preset to switch at 4.5 amperes, an overload of 4.4 to 4.6 amperes causes the protective circuit to remove the primary voltage from the regulated supplies in about 0.020 second. If required, still better precision can be realized where the tolerances on the variations of supply voltage and frequency are smaller, or where compensating features are added. Up to about 97% of the supply voltage can be suppressed easily, while the normal drop at full output is about 3%.

General Circuit Operation

The block diagram shown in Fig. 1 represents one type of magnetic-amplifier protection circuit which has been de-

---

Paper 57-771, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 21, 1957; made available for printing June 11, 1957.

R. J. DISTLER and C. STURZENBECKER are with Bell Telephone Laboratories, Inc., Whippany, N. J.

Vo
Pt.
19

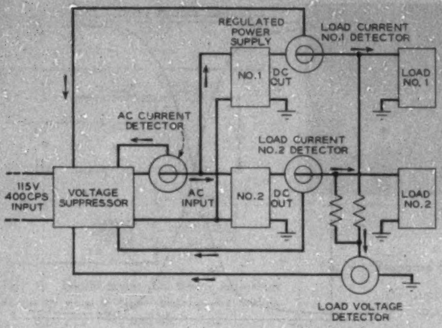


Fig. 1. Current-voltage overload protection block diagram

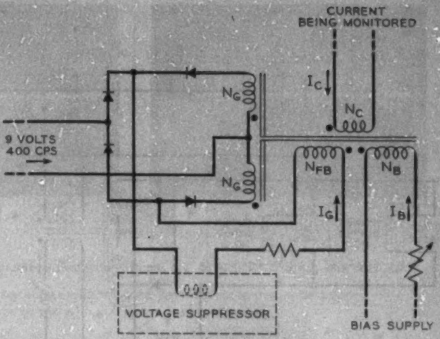


Fig. 2. Detector circuit

veloped. Power supplies 1 and 2 are each protected against excessive output current, output voltage, and input current. One output lead of each power supply is wired through the center hole of its toroidal current detector assembly, effectively forming a single turn on the toroid. A single voltage detector uses resistor summing to monitor the two output voltages. Separate voltage detectors may be utilized if desired, and the summing is then done at the input to the voltage suppressor. One of the a-c power leads common to both supplies is wired through the center hole of its toroidal a-c current detector assembly.

If any one of the detectors monitors an overload, that detector will forward a signal to the voltage suppressor, which will in turn respond by rapidly absorbing the input voltage. Since the voltage suppressor is in series with the power supplies, the voltage to the power supplies is reduced to a value where the supply circuits are safe. As a result, the loads and power supplies are protected against damage.

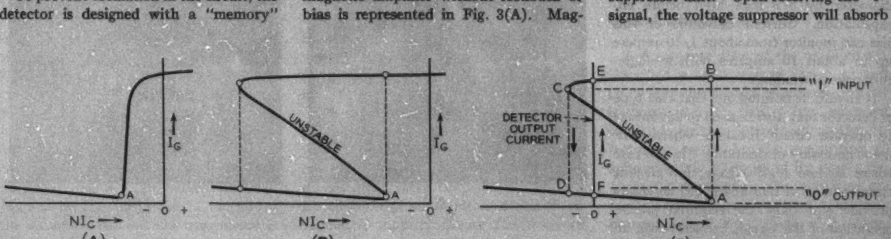
To prevent oscillation in the circuit, the detector is designed with a "memory" which causes the circuit to continue its signal to the voltage suppressor until a reset operation has been performed. Many supplies can be protected with one voltage suppressor unit by inserting the proper detectors in the circuits where protection is desired.

Current and Voltage Detector Operation

The detector circuits are essentially bistate magnetic amplifiers as shown in Fig. 2. Each detector assembly uses two toroidal cores. Winding N_G is wound on each core, then the two cores are stacked and windings N_{FB}, N_B and N_C are wound common to both cores. In the current detector a 1-turn control winding N_C is formed by passing the power supply load lead through the toroid. In the voltage detector, the N_C winding consists of a number of turns in order to make it possible to monitor a small amount of current proportional to the voltage.

The gate current I_G of a self-saturating magnetic amplifier without feedback or bias is represented in Fig. 3(A). Magnetic positive feedback is obtained in the circuit of Fig. 2 by connecting the feedback winding N_{FB} in series with the gate circuit so that the gate current I_G flows through the feedback winding. This feedback causes the transfer characteristic of Fig. 3(A) to bend back at point A, Fig. 3(B) and form an S-shaped curve. By applying a direct current through the bias winding N_B of Fig. 2, the transfer characteristic is shifted to the right by an amount equal to the ampere turns of the bias winding. The transfer characteristic shown in Fig. 3(C) is the result.

The current I_C in the winding N_C is the load current of the power supply being monitored in the case of the current detector, or a current proportional to the voltage being monitored in the case of the voltage detector. When this current, which can be either alternating or direct, increases to a point which is to the right of point A, Fig. 3(C), and is retained for the necessary switching time, detector output current I_G snaps to the level at point B. The "1" or full detector output provides a "1" signal to the power voltage suppressor unit. Upon receiving the "1" signal, the voltage suppressor will absorb



(A) (B) (C)

Fig. 3. (A)—Transfer characteristics without feedback or bias. (B)—Characteristic with feedback, no bias. (C)—Characteristic with feedback and correct bias

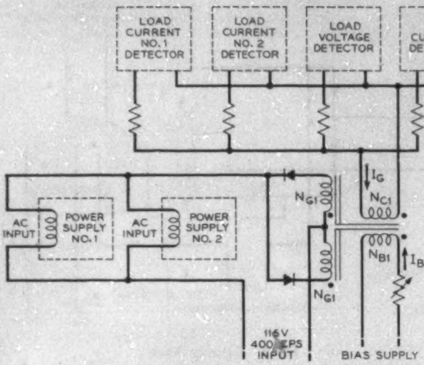


Fig. 4. Voltage suppressor circuit

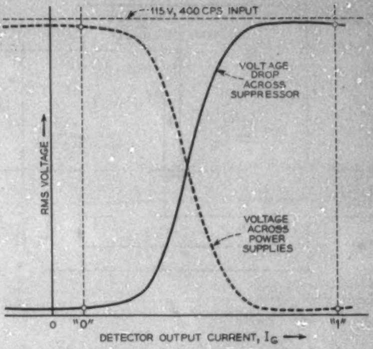


Fig. 5. Voltage versus I_G for voltage suppressor

nearly all the primary voltage and the outputs of the power supplies will drop to nearly zero. Concomitantly, I_C drops to point E, but the detector output remains at the higher level (point E), and the "1" detector output is retained—this is the memory function which has been mentioned. To bring the current of the detector back to the lower level of I_G, or "0" condition, the bias current in winding N_B must be increased to shift the characteristic until point C is to the right of $NI_C=0$. At this point I_G will snap back to its lower level, point D, and return to point F when the bias is returned to normal. Alternatively, the gate current can be interrupted to reset the detector to 0. This same type of circuit could be used to detect an "under-current" or "under-voltage" if this were desired. If point C is maintained to the right of $I_C=0$, the circuit will oscillate around the $DABCD$ loop whenever there occurs an overload.

Each detector unit of the experimental model measures approximately 1 inch in diameter by 5/8 inch high before encapsulation. Current detectors of this size can monitor from about 1/10 ampere up to about 10 amperes with a single-turn control winding.

It should be pointed out that this type of detector may also be used to advantage to operate circuit breakers where their use is necessary or desirable. In one case where several high-voltage, low-current circuits are to be monitored, the detectors provide not only the required precise operation of the circuit breaker when an abnormal condition is monitored, but also make possible the electrical isolation and remote mounting of the circuit breaker with respect to the high-voltage cables. When an abnormal condition is monitored by one of the detectors, the circuit breaker operates. This removes the monitored high voltages from the circuits until the breaker is manually reset.

In this case, the only function of the detector memory characteristic is to maintain a voltage across a lamp indicator designating the detector which monitored the abnormal condition.

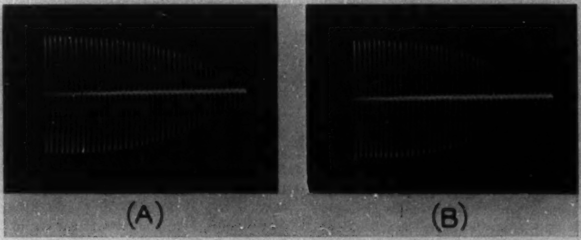

(A) (B)

Fig. 6. Response time of basic protection circuit

A—Suppressor output with overload just sufficient to trigger the detector
B—Suppressor output with 200% overload monitored at the detector

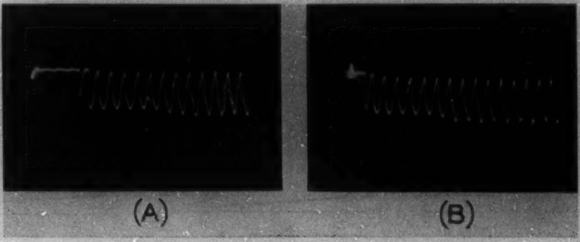

(A) (B)

Fig. 7. Detector output showing the response of detector alone

A—Just sufficient overload to trigger the detector
B—200% overload

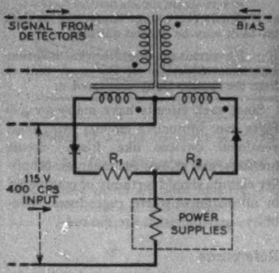


Fig. 8. Modification of basic suppressor circuit to provide faster response by adding R1 and R2

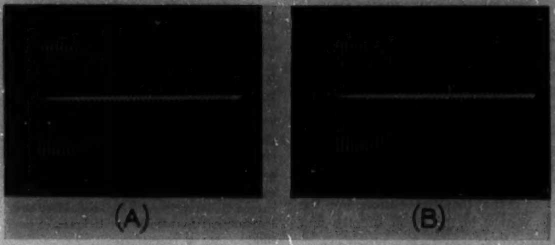


Fig. 9. Response of modified protection circuit, Fig. 8, with R1 and R2, one ohm each

A—Suppressor output with overload just sufficient to trigger detector
B—Suppressor output with 200% overload at detector

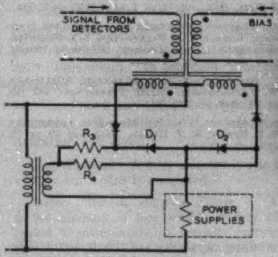


Fig. 10. Modification of basic suppressor circuit to provide faster response by adding biased diodes D1 and D2

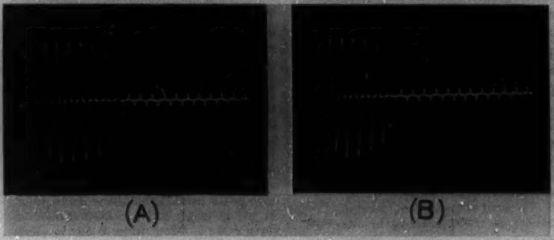


Fig. 11. Response of modified protection circuit, Fig. 10

A—Suppressor output with overload just sufficient to trigger detector
B—Suppressor output with 200% overload

Voltage Suppressor Operation

The function of the voltage suppressor is to absorb the primary voltage (suppress it from the power supplies) when a "1" signal is received from any one of the detectors. Since the voltage suppressor is connected in series with the power supply inputs, the voltage absorbed by the suppressor circuit should be negligible when a "0" signal is received and large when a "1" signal is received. The voltage drop in the model tested is about 3% of line voltage in the "0" state and about 97% of line voltage in the "1" state.

Basically, the circuit is a conventional doubler magnetic amplifier as shown in Fig. 4. The characteristic of the control winding current I_G versus the alternating voltage drop across the suppressor is shown in Fig. 5. The current through N_{C1} is either the "0" or the "1" output of the detectors so that the suppressor is either in the "0" or "1" state and does not operate in its proportional region.

The desired $I_G=0$ point is obtained by adjusting the current in the bias winding N_{B1}. At the "0" operating point of I_G, most of the primary voltage is across the power supplies. However, at the "1" point of operation, most of the primary voltage is absorbed by the voltage suppressor N_G windings, and the power supplies are essentially shut off.

The voltage suppressor was designed to pass about 5 amperes and suppress 115 volts at 400 cps; before encapsulation it measures 3 inches in diameter by 3 inches high. If removing, say, 50% of the alternating voltage to the power supplies would be sufficient, the voltage suppressor could be reduced in size accordingly.

One circuit parameter which becomes of particular importance in the design of the voltage suppressor circuit is the surge current capacity of the diodes. These diodes must carry the fault current referred to the power input terminals of the regulated supplies during the response time of the detector and voltage suppressor. Hence, the power that a specific set of diodes will handle may be increased by reducing the over-all response time or limiting the fault current. Of course, this factor may be eliminated from consideration by using a saturable reactor, which has no diodes. However, such a circuit gives a considerably long response time, and thus poorer protection, as well as requiring larger magnetic components. The greatest improvement may be realized by reducing the response time.

Over-All Response Time

The response time of the basic circuit described is illustrated in Fig. 6(A) and (B) where the voltage suppressor output is shown as a function of time when an overload signal is introduced at the detector at $t=0$. A time amounting to approximately 30 cycles of the power frequency is required for the combined detector and voltage suppressor operation; however, only 1/2 to 2 cycles of this time, depending on the degree of overload, is required for the detector response, as is shown in Fig. 7(A) and (B) where the detector output has been illustrated.

One method of reducing this response time which at the same time limits the suppressor output current is to insert a small-value resistor in series with each diode as shown in Fig. 8. As is shown in Fig. 9(A) and (B) the response time of the protection circuit is reduced by approximately one half when $R1$ and $R2$ are each one ohm.

There is a disadvantage in this method which is that the series resistors drop the line voltage to the power supplies and waste power.

A better method of acquiring faster response is shown in Fig. 10 where the diodes $D1$ and $D2$ are biased from a separate source. This configuration also limits the current due to a fault since any current greater than the biased diode current must flow through $R3$ and $R4$. The line voltage is not dropped by series resistance in this case and the power dissipation is relatively small. Yet, the response time, Fig. 11(A) and (B) is reduced to about one fifth that of the basic circuit, or to about 15 milliseconds when 400-cps power is used.

Conclusions

The circuit described fulfills the protection needs for low-voltage high-current power supplies. Some advantages are:

1. The circuit contains no moving parts or electrical contacts such as are found in circuit breakers and relays.
2. The circuit contains only magnetic cores with windings and diodes; therefore rugged construction and high reliability over extreme environmental conditions are possible.
3. The protection circuit does not depend on heating or magnetic pull for operation.
4. Operation of the circuit in about 7 cycles of the supply frequency is possible.
5. Electric arcs which are experienced when a high-current circuit is broken with contact devices are eliminated.
6. The current detector introduces no voltage drop in the circuit being monitored.
7. The point at which overload protection occurs may be easily and precisely adjusted.
8. The circuit will monitor alternating or direct current or a combination of the two.
9. Each detector can operate an indicator as it signals the voltage suppressor so that the fault may be located quickly.
10. The circuit provides electrical isolation and remote location of the reset control from the circuits being monitored.

Since such circuits have many advantages over protective circuits using conventional devices like fuses, circuit breakers, and relays, the magnetic-amplifier circuits should certainly be considered in all instances where ruggedness, reliability, and precision are desired.

References

1. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955.
2. Magnetic Amplifiers in Bistable Operation, L. A. Finzi, G. C. Feth. *AIEE Transactions*, vol. 74, pt. I, Nov. 1955, pp. 592–98.
3. A Magnetic Amplifier Switching Matrix, D. Katz. *Ibid.*, vol. 75, pt. I, May 1956, pp. 236–41.
4. Transient Processes in Contactless Magnetic Relays, I. B. Negnevitzkyi, L. L. Samurina. (translation from Russian). *Engineers Digest*, New York, N. Y., Apr. 1954.
5. Self-Saturation in Magnetic Amplifiers, W. J. Dornhoefer. *AIEE Transactions*, vol. 68, pt. II, 1949, pp. 835–50.
6. Response Time of Magnetic Amplifiers, E. L. Harder, W. F. Horton. *Ibid.*, vol. 69, pt. II, 1950, pp. 1130–41.

Design Factors for the Optimization of Multichannel Radio Systems

CHARLES A. PARRY
NONMEMBER AIEE

Synopsis: The satisfactory design of a multichannel radio system requires evaluation of both nonlinear and thermal noise power and the statistical considerations associated with each. It is pointed out that an optimum design can be attained when these noise powers are equal; and since the noise power is greatest in the top channel of the FM (frequency modulation) system, it is sufficient and suitable to establish design criteria based on these premises. It is seen that certain essential performance factors are required for this purpose. These include a measurement of signal/noise ratio using a white-noise test signal, a knowledge of the principal orders of distortion which give rise to the nonlinear noise, and a value for the overload power of the base-band signal.

The effect of such factors as carrier power, channel spacing, speech-power distribution, and traffic activity, on the signal/noise ratio is then examined. The basic expressions are also used to establish the total reliability as a function of traffic fluctuations, propagation reliability, and carrier power. Theoretically optimum distributions as well as typically operative distributions are given.

In general, systems are of two classes: integrated—those which must conform to the technical requirements of existing or proposed associated networks, and nonintegrated systems—those which exist, as it were, independently. Standards are suggested for both these systems. The CCIF (Comite Consultatif International Telephonique) fictitious circuit of reference is a special case of the former and it is concluded that the signal/noise ratio suggested for this reference is some 10 db (decibels) too high. System testing by means of multiple tones is also discussed and some basic requirements related to this method are outlined.

ONE of the principal technical requirements in the design of a multichannel radio system is to establish a satisfactorily low level of noise in the output of the various channels. This noise may be separated into two components: thermal and nonlinear. Thermal noise is related to those factors associated with transmission loss, while nonlinear noise, which arises from distortion of the multichannel signal, is related to the traffic handling capabilities of the system. The over-all performance can, therefore, be expressed as a function of system parameters which utilize these relations; also, it is possible to determine those conditions for which the total noise is a minimum, thereby defining an optimal behavior.

This provides a reference design framework which is useful for studying the economics of system planning and for establishing realistic performance standards and objectives. In addition, it is frequently useful in preliminary commercial negotiations since the concept of decibels (of signal/noise ratio) per dollar is readily established. This paper is concerned with these basic relationships.

In practice, the total noise is due to contributions from both radio and multiplex equipments. Ordinarily, this latter is small and only the contribution from the radio equipment is considered here. All systems referred to are those which use single side-band suppressed carrier frequency division multiplex equipments with FM radio transmissions.

Paper 57-60, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted August 27, 1956; made available for printing May 17, 1957.

Charles A. Parry is a consultant with Page Communications Engineers, Inc., Washington, D. C.

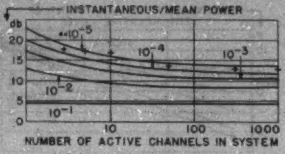


Fig. 1. The multichannel peak factor

++ These points are taken from Fig. 4 of Holbrook and Dixon's paper
—— Taken from Jacobsen's paper

Definitions

Although many of the terms which must be employed in a study of this kind have appeared from time to time in the relevant literature, it is necessary to redefine some of them so that their implication in the present context should be fully understood. In addition, new parameters are needed to avoid unnecessary confusion and these also require explanation. Accordingly, the following definitions are given.

Busy Channel. A busy channel is one to which a subscriber is connected.

Active Channel, Activity Factor τ. An active channel is one which, during the period of observation, is actually carrying speech power. This period includes the short pauses associated with normal speech. The activity factor is the fraction of time for which a channel is expected to be active. Measurements[1] indicate that this has a maximum value during the busiest hour in a large exchange of $\tau = 0.25$.

Channel Capacity. This is the maximum number of equispaced channels which can be used with a given equipment.

Fully Loaded System. A fully loaded system of N_r channels is one which has N_r busy channels.

Partially Loaded System. A system with N_r channel capacity is not necessarily always used in a situation calling for N_r available channels. The number N which is actually required may in fact be considerably less than the number N_r. A system so used is a partially loaded system. This is not the same as an N_r channel system which does not have N_r channels busy all the time.

Optimised System. An optimized system as discussed in this text is one in which the system parameters are chosen so as to produce the maximum possible signal/noise ratio in some specific channel for a specific set of conditions.

Reference Mean Speech Power. This is taken to be the mean power of the single talker volume distribution. In the present text, this is taken as equivalent to -12.5 dbm (decibels referred to 1 milliwatt) measured at a point of zero relative level.

Multichannel Peak Factor. This is used to express the ratio between the instantaneous peak voltage and the rms (root-mean-square) voltage of the complex wave due to the composite signal from a number of talkers. It should be noted that this factor is associated with a probability distribution. Curves relating to this are given in Fig. 1 and are further discussed in the text.

Overload Expectancy ϵ. This is defined as the probability that a given multichannel peak factor will be exceeded.

Equivalent Power of the Multichannel Signal. This is the average power of the combined signal of the several channels. It is further discussed under this section in the text.

Excess Power Probability p. The equivalent power of a multichannel signal is determined on a statistical basis. Hence, for a given set of conditions there is a definite probability that a specific power will be exceeded. This may be defined as the excess power probability p. It is convenient to express this as the percentage of time for which the power will exceed that indicated.

Overload Point (of a Device). That point on the input output amplitude characteristic where a linear relationship no longer exists. It may be defined in two ways. The CCIF[2] define this point as that for which an increase of 1 db in fundamental signal produces a 20 db increase in the third harmonic. Hence, it may be defined by that power for which the distortion ratio departs by some specified amount from the value which would be expected if the overload did not occur. The region in the vicinity of the overload point is not sharply defined in practical equipment, but exhibits a gradual transition from linearity; the transition region extending over an input range of perhaps ±1.5 db. In systems where many devices are in tandem, it is sometimes necessary to define the overload point as that corresponding to some specific departure from linearity, since this cumulative departure may be a determining factor.

Instantaneous Load Capacity, Load-Capacity Factor. The instantaneous load capacity of an equipment is a peak power requirement necessary to ensure that interference from overload effects will be negligible. (This interference may be referred to as "overload distortion" and is generally due to 3rd and higher order products.) As originally established by Holbrook and Dixon,[1] this is finally expressed as the mean power of a sine wave with a peak power equal to that of the complex signal, relative to a 1-mw (milliwatt) test tone, measured at a point of zero relative level. The curve so established by these authors is shown in Fig. 2. It is determined in such a way that if the overload power of the device is made to coincide with this peak power, then those peaks of voice energy which are in excess of this and which, therefore, lie in the overload region of the equipment, will cause negligible interference in the channels of the multichannel system. The ratio, mean power in a sine wave with peak power equal to that of the complex signal, to single-channel test tone, may be regarded as the load-capacity factor. If the load capacity factor is used with respect to the overload point, then the maximum permissible transmission level of each channel can be established.

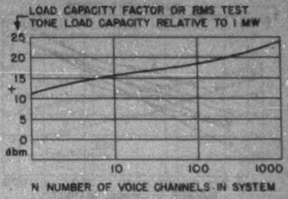


Fig. 2. Instantaneous load capacity, with peak limiting on all channels and with uncontrolled volumes. After Holbrook and Dixon

$p = 0.01$ $\tau = 0.25$ $\epsilon = 10^{-5}$

Nomenclature

X, Z = proportionality constants
τ = activity factor
α, a = exponents associated with the relation between single channel test tone and equivalent power or peak power, respectively
A = ratio between accepted mean power of single talker volume distribution and -12.5 dbm
P, P_c, P_p = power ratio referred to single-channel test tone. In the order given, the ratios are: equivalent power to tone; tone to reference power; peak power to tone
P_m = ratio of overload power to reference power
N = number of contiguous channels in system
N_r = total number of contiguous channels of which the equipment is capable. Test data is often expressed in terms of N_r rather then N
S = signal/noise ratio
S_n = signal/thermal noise
S_e = signal/nonlinear noise
S_t = signal/total noise. This symbol is reserved for the maximum possible ratio; the noise is weighted
K = noise weighting factor employed on voice channels
b = band spacing between channels
W_0 = median carrier power at receiver input
F_0 = effective noise figure of receiver
r_0 = factor related to fading margin and the number of hops
ψ_n, ψ_c = ratio of thermal or nonlinear noise power to single-channel test-tone power, respectively. (The noise is weighted.)
f = any frequency in the base band
f_m = highest frequency used in the base band
d_r = reference deviation at which white-noise test data is obtained
γ_N, γ_r = signal/noise ratio measured in an N or N_r channel system, respectively. This figure is unweighted.
ϕ, $(f/f_m)^\phi$ = ratio of distortion powers at frequencies f and f_m, respectively
δ = factor by which the base-band signal power is increased in order that the thermal and nonlinear noise powers in the measured channel should be equal
β = order of distortion giving rise to nonlinear noise
ϑ = factor by which the signal/noise ratio in

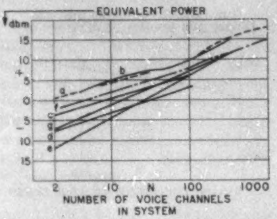


Fig. 3. Equivalent power of N-channel system taken from various authors

a—After Jacobsen, "Combined mean power of R subscribers speaking half the time"

b—From Brockbank and Wass, Fig. 6, "Average distortion curve"

c—After Slow, "Root mean square level exceeded 1% of the busiest hour"

d—After Brown, "Speech power exceeded 1% of the time"

e—After Lewin, assumed mean power for single talker −15 dbm

f—After Holbrook and Dixon, "Equivalent volume exceeded 1% of the time with uncontrolled volumes"

g—After Holbrook and Dixon, "Equivalent volume exceeded 1% of the time with controlled volumes"

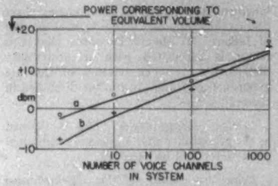


Fig. 4. Equivalent power of multichannel signal

a—From Holbrook and Dixon, "Equivalent volume exceeded 1% of the time"

b—From Holbrook and Dixon's paper, "Equivalent volume exceeded 10% of the time"

o—From Jacobsen's Fig. 1(B), using 1% probability curve for half the subscribers speaking half the time

++—Also from Jacobsen, using 10% probability curve

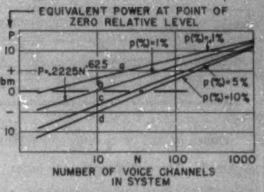


Fig. 5. Equivalent power of multichannel signal during busiest hour, with activity factor $\tau=0.25$. Curves based on uncontrolled volumes and nonlimited channels

the top channel is increased due to the use of base-band equalization, the equivalent power of the complex signal being unchanged in this process.

The Equivalent Power of the Multichannel Signal

In order to establish the equivalent power of a multichannel signal under conditions of actual operation, relative to that of a single channel test tone, a detailed statistical study of speech distribution and voice circuit selection is necessary. A preliminary study of the relevant literature would seem to indicate a certain amount of confusion has existed. Fig. 3, for instance, shows curves relating to this subject taken from a number of authors. A more detailed examination, however, tends to clarify the picture.

Brown[3] and Slow[4] published curves based on contemporary authors. Lewin[5] established a relation which did not take into account the probability distribution of the active channels. Although useful for a theoretical investigation, it is not related to practical conditions and should be discarded in optimum designs. Brockbank and Wass[6] established a curve based on what they term "average distortion" but the exact derivation is not clear. Holbrook and Dixon[1] probably paved the way in the studies on this subject. Their detailed mathematical analysis is backed up by intensive experimental work under practical conditions and their data must be regarded as being as reliable as the state of the art will allow.

Jacobsen[7] also made a very fine analysis in considerably more detail than Holbrook and Dixon. If we apply the same activity factor to the two sets of data, the results are in good agreement as shown by Fig. 4. This encourages the more extensive use of Jacobsen's data to provide detail not readily possible with that of Holbrook and Dixon. The present discussion will be confined to the data available from these two papers.

One final difficulty remains. The unit corresponding to the mean power of the single talker distribution was set by Holbrook and Dixon at about −10 dbm. Brockbank and Wass indicated that measurements by the British post office showed a value some 5 db below this. Subsequent investigation[8] since the original computation of Holbrook and Dixon's data would seem to indicate their original value was somewhat too high. A mean value of −12.5 dbm would appear to be a satisfactory figure for present day designs and is recommended as a value for reference mean power (equivalent power rather than equivalent volume has been used throughout this paper for convenience). This means a −2.5 db correction figure has to be applied to the equivalent power derived from their data. Using this correction, it is possible to draw up a series of curves for various excess power probabilities with an activity factor 0.25, corresponding to the factor observed at the busiest time in a large trunk exchange. Such curves are given in Fig. 5.

Fig. 6 gives similar curves obtained for an activity factor of 0.125, which might be defined as half as active as the greatest activity of the channels in a large trunk exchange. In drawing up this data, it is found that the curvature obtained from the direct use of the points is very slight so that the nearest possible straight lines have been used.

It can be shown that the effect of limiting in each channel tends to reduce the equivalent power slightly.[1,7] However, the curves of Figs. 5 and 6 have been derived without limiting, since it is felt that the reduction in power which will occur, represents a suitable design margin which will assimilate the slight increase due to carrier leak, signalling tones and the like. (This is generally true when p is less than about 0.05, which is the usual range of interest. However, if the mean power of these signals is at all appreciable, the distributions shown in Figs. 5 and 6 will be affected, particularly for the larger values of p. For this reason, it is always advisable to check the relative power of these signals.) In addition, they are obtained for uncontrolled volumes since this is the condition which pertains in a practical, operative system.

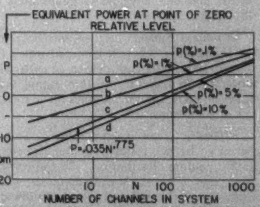


Fig. 6. Equivalent power of multichannel signal during busiest hour, with activity factor $\tau=0.125$. Otherwise similar to Fig. 5

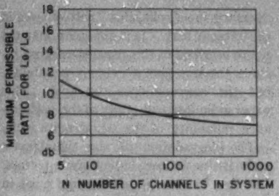


Fig. 7. Minimum relation between the overload power L_o and the mean power L_a of the multichannel signal

Examination of these curves brings to light a most useful relation. It is seen that in general all these curves obey the law

$$P = XN^{\alpha} \tag{1}$$

where P is the ratio between the mean power of the composite signal and that of the 1-mw test tone, N is the number of channels, and X is a constant of proportionality. The value of P is the value required in the computation of nonlinear noise in the system.

It is possible that some administrations might not find acceptable the use of the reference mean speech power of −12.5 dbm recommended herein. In order to overcome this difficulty in subsequent analysis, it is proposed to introduce a factor A which could take any differences into account. The factor A may be defined as the power ratio between the expected value of the mean power of single-talker volume distribution, and a power of −12.5 dbm. For instance, if the figure of −15.0 dbm as suggested by Brockbank and Wass is used, A would be the ratio corresponding to −2.5 db or $A = 0.56$. Hence, the law of equation 1 is modified to

$$P = XAN^{\alpha} \tag{2}$$

The curves of Figs. 5 and 6 are based on a recommended reference power of −12.5 dbm so if these curves are used, $A = 1$. If p remains fixed, this law may also be expressed in another useful form,

$$P = XAN^{\alpha}\frac{\tau^{\alpha}}{0.25}$$

It should be observed that the ratio between the expected power of the composite signal and the single-channel test tone cannot be defined arbitrarily because it is associated with a probability distribution. Hence when speaking of the equivalent power, it is also necessary to define the probability and activity factors with which it is associated.

It is also important to note that the equivalent number of active channels is not τN as is commonly used in some of the literature. Noise analyses based on this latter assumption are not correctly related to the behavior that might be anticipated in typical operative systems. This has particular significance when white-noise testing is used; a suitable correction factor must then be introduced.

The Load Capacity Factor

The load-capacity factor is determined from the distribution of the instantaneous peak power of the complex multichannel signal. Fig. 2 shows the curve established by Holbrook and Dixon expressing this factor as a function of the number of channels in the system.

In order to derive a relation of this kind, it is necessary to determine the multichannel peak factor and equivalent power of the composite signal. These have been investigated separately by Holbrook and Dixon and by Jacobsen. The results of these authors are found to be in good agreement.

Peak distribution curves derived from these sources are shown in Fig. 1. Equivalent power distribution curves are given in Figs. 5 and 6 for activity factors $\tau = 0.25$ and $\tau = 0.125$, respectively. The proper addition of suitable peak and equivalent power distribution curves gives a value for the instantaneous peak power. However, although the equivalent power maintains a discrete value long enough to contain a representative sample of the multichannel peak factor distribution, it is, in effect, a constantly changing parameter and the addition so obtained can be expected to apply only for a small part of the total operative time. Consequently, the total probability for this instantaneous peak power will be considerably less than that of either factor. When the effect of limiting in each channel is included, this probability is still further reduced, particularly when the number of channels is less than about 12. The curve of Fig. 2 was established from the separate curves for $\tau = 0.25$, $p = 10^{-2}$, and $\epsilon = 10^{-5}$, with an allowance for limiting. It can be expected, therefore, that peaks in excess of that shown by this curve will occur for considerably less than 0.001% of the time, and a value of perhaps 10^{-4} or 10^{-5} % is more representative.

The effects of overload vary from equipment to equipment and with the number of units in tandem and a certain amount of margin also has to be allowed for component variations. However, the extremely small probability of overload indicated by Fig. 2, even with these considerations, does not lend itself to the most economical design.

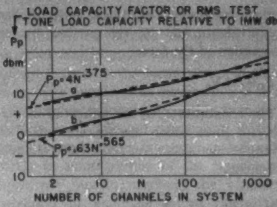


Fig. 8. The instantaneous load capacity P_p with peak-limited channels and uncontrolled volumes

a—Suggested curve for integrated systems
b—Suggested curve for nonintegrated systems

In a practical multichannel radio system, the effective nonlinear noise will be related directly to the same equivalent power curve as is used to obtain the load capacity factor. This means, for instance, in the above case, that this nonlinear noise can be expected to exceed some specified value for 1% of the time during which all the channels are busy. This is also the period when it may be expected that overload would be most likely to occur. Hence, the probability of overload need only be substantially less than the probability of excess nonlinear noise. The overload effects then may be expected to be lost in the nonlinear noise. While it is true that the choice of overload expectancy may be varied by empirical considerations, a value of $\epsilon = 0.001$, for instance, is still only one-tenth the probability associated with the equivalent power (corresponding to $p = 0.01$, $\tau = 0.25$) referred to above. It can be expected that the resultant instantaneous peak probability over the full working period will be substantially below this, so that, for practical systems, it should be quite suitable to establish curves based on $\epsilon = 0.001$.

As stated earlier the original analysis assumed a reference mean speech power of −10 dbm while the present data are based on −12.5 dbm, so a −2.5 db correction must be made. Fig. 7 shows curves made with this correction and an allowance for limiting, based on $\epsilon = 0.001$.

It will be seen that these curves can be very nearly represented by a straight line. The use of such an artifice is justified because the actual peak requirement is not absolutely defined, but is the nearest assessment of the data. The small margins obtained with the use of straight line graphs are quite acceptable in practical systems. As in the case of the equivalent power curves, these straight line equivalents can be represented by an equation of the form

$$P_p = ZAN_a \qquad (3)$$

The constants Z, a, are indicated in Fig. 8, for lines which are representative of requirements between 2 and 1,000 channel systems.

Categories of Service

It would seem that there are two separate categories of service which must be considered in the design of multichannel radio links. These have been tentatively identified by the terms "integrated" and "nonintegrated" systems. The first of these is concerned with those links which must be interconnected with existing communication facilities established on a national or quasi-national scale. The need for maintaining a certain standard of service in each link of such a network establishes the performance requirements for each additional link. Multichannel radio systems associated with such networks would be expected to give toll quality performance, conforming generally overseas to standards recommended by the CCIF. Systems in this category might be expected to be fully loaded so that in an N_r system there would be N_r busy channels during the busiest period of service. Therefore, the design would need to be based on an activity factor τ of 0.25. The equivalent power would be expected to be given satisfactorily by Fig. 5, curve b, which can be represented by the equation $P = 0.2225N^{0.625}$. The corresponding load capacity factor Fig. 8, curve a, represented by the equation $P_p = 4N^{0.375}$. The CCIF fictitious circuit of reference can be regarded as a particular type of integrated system.

The second category of service is determined by those systems which do not have to be interconnected with surrounding networks, except perhaps purely on a temporary basis or during a period when some special emergency exists. The operative standards of such systems may, therefore, be defined in terms of minimum performance requirements, consistent with good engineering practice and with associated economic factors.

Usually the channels of these systems are not as busy as those of the former type. The channels provide the facility of contact between points rather than being continually in use. For this reason, the activity factor is unlikely to be as high as that of the fully loaded system. The level of activity which should be ascribed to such systems is a matter of judgment. It would seem that an activity half that of a continually busy channel would be a reasonable one. The value of τ for this is 0.125.

In the same way, it is not necessary to consider an excess power probability as small as that suggested for the integrated system. The value used in the present text is 5%, corresponding to a traffic reliability of 95%. This matter is also discussed later. The equivalent power for this figure is given by curve C of Fig. 6 for which the law is $P = 0.035N^{0.775}$. The load capacity curve suggested for this category of service is given by curve B in Fig. 8 represented by the equation $P_p = 0.63N^{0.565}$. Typically, the number of channels associated with this type of system is relatively small, the maximum being probably in the order of 24. Once the number exceeds this, it is not uncommon to find that there are associated factors operative which demand the integrated system approach.

Reliability

There is little doubt among engineers that there are two connotations for the word "reliability," and a note on the subject seems on order here. The most fundamental concept is that the term applies to a "go or no go" situation, and, as such, it applies to system outage. This will be due either to abnormal propagation conditions or equipment failure. In this sense, the understanding of reliability is the "probability of survival." For instance, for a failure rate of 0.1%, the probability of survival is 99.9%. If the word reliability is retained, a preferable term, perhaps, would be "circuit reliability."

On the other hand, it is quite common to accept the postulate in negotiation that the equipment failure rate is very small. In this case, reliability is used to indicate that a given grade of service will be maintained. This is expressed by the stating that the signal/noise ratio measured in the channels will not be worse than some specified value for a given percentage of time. In this sense, the use of the word reliability tends to be confusing to all but the engineer; because it is not implied that the system is inoperative for the remaining percentage of time, but rather that the service will be degraded. Quite often this degradation is comparatively small and the point has an academic value rather than a practical one. The understanding of reliability here is quality probability, i.e., the probability that a given circuit quality will be maintained. To clarify this usage, the generic term "performance* reliability" will be used in the present text. It is apparent that the figure ascribed to this parameter must include the loss due to circuit reliability, but the converse does not hold. This paper is limited to a study of the performance reliability of the signal/noise ratio.

The degradation of the signal/noise ratio is due to two independent factors. The first of these is the thermal noise in the system, which is affected by propagation. The term "propagation reliability" is used in connection therewith. The second factor is the nonlinear noise, sometimes called crosstalk. This is determined by the traffic on the channels of the system and the term "traffic reliability" has been suggested. This term will be used in the present paper. It will be seen that the traffic reliability factor is $(1-p)$ where p is the excess power probability. When stating this reliability, it is necessary to stipulate the activity factor τ with which it is associated, otherwise the term has no physical interpretation.

It should be noted that traffic reliability usually pertains to the busiest hour, which itself represents only 4.1% of the total time. On the other hand, propagation reliability usually relates to much longer periods. Care must be taken to ensure that percentage-of-time figures refer to the same base period otherwise total-time figures will be in error.

Thermal Noise

The manner in which the thermal noise falling into the channels of a multichannel system may be calculated has now been treated extensively in the literature and will not be detailed here. The basic expression for the signal/weighted noise ratio measured in a bandwidth b cycles wide at the output of the FM receiver may be expressed in the following way,

$$S = \frac{W(d_c^2)3}{2kTFK[(f+b/2)^3-(f-b/2)^3]} \qquad (4)$$

where

W = r-f (radio frequency) carrier power, assumed considerably greater than the equivalent noise power at the receiver input
d_c = peak deviation in cycles produced by the signal in question
k = Boltzman's constant
F = noise figure
f = mid-frequency in cycles of the band being measured
K = weighting factor = 1, if the unweighted ratio is required
T = absolute temperature

---

* The term "degradation time" is also a useful concept here. This may be defined as the period of time based on the reliability data, for which the stated service is actually degraded.

Obviously, as is well known, S has its minimum value in the top channel of the multichannel system and for this we may state that $b<<f$. In practice, for the voice channels ordinarily met with, Kb is virtually constant with the value $1.5\times(10^{8})$.

Effecting the simplifications and numerical substitutions gives for the top channel in the system

$$S_1=\frac{Wd_c^2 8\cdot4(10^{16})}{F(f_m)^2} \qquad (5)$$

where f_m = highest base-band frequency.

If the deviation is that which corresponds to a single-channel test tone, then of course S_1 is the test tone/weighted thermal noise ratio in the top channel.

Equation 5 gives the ratio obtained from only one receiver. In the over-all link there may be many such receivers and the noise powers due to each will add in the channel. Therefore to evaluate the over-all effect it is more convenient to use the noise/signal ratio. We can then express this over-all value as follows

$$\psi_n=\left(\frac{f_m}{d_c}\right)^2\frac{1}{8\cdot4(10^{16})}\left(\frac{F_0r_0}{W_0}\right)$$

$$r_0=\frac{W_0}{F_0}\left(\frac{F_1h_1}{W_1}+\frac{F_2h_2}{W_2}+\dots\frac{F_Hh_H}{W_H}\right) \qquad (6)$$

where

ψ_n = weighted thermal noise/signal power ratio in the top channel of the link consisting of H hops

$W_{1,2},\dots$ = carrier power inputs to the 1st, 2nd, etc., receivers

$h_{1,2}\dots$ = fading margins required for the respective hops. (The margin is here expressed as a factor rather than in decibels)

$F_{1,2},\dots$ = effective noise figures for the receivers

W_0, F_0 = normalizing figures corresponding to median values of W and F at any receiver

The particular value of 6 is that the influence of variations in any W or F over any part of the link may be readily determined. It is apparent that is a value is assigned to ψ_n in equation 6 then a specific requirement for W_0/F_0r_0 is established.

Usually it is considered that fading occurs effectively over only one hop at a time; the amplitude of the fade such that contributions to the noise from the $(H-1)$ receivers can be neglected. If the noise figures $(F_{1,2},\dots)$ are substantially equal then we can put $r_0=h_1$. Hence the transmission requirements for each hop can be obtained from the relation

$$\frac{W_0}{F_0r_0}=\frac{W_1}{F_0h_1}=\frac{W_2}{F_0h_2}=\dots$$

Again, when the fading margin is comparatively small, it may be assumed that the $(H-1)$ hops cause noise which is at least equal to half this number with the carrier at median level. Then we would have

$$r_0=h+(H-1)/2$$

and the r-f transmission requirements for each hop would be obtained from the relations

$$\frac{W_0}{F_0r_0}=\frac{W_1}{F_0\left[h_1+\frac{H-1}{2}\right]}=\frac{W_2}{F_0\left[h_2+\frac{H-1}{2}\right]}$$

When the noise figures are not similar, or complicated fading patterns are expected, then of course different relations can be established to determine the transmission details.

If d_a is the derviation corresponding to the mean power of the composite signal then

$d_c^2P=d_a^2$ and

$$\psi_n=\frac{F_0r_0Pf_m^2}{W_0d_a^2(8\cdot4)(10^{16})} \qquad (7)$$

The reciprocal of this, S_n, is of course the over-all signal/noise ratio in the channel.

In practice, it is convenient to refer to some reference deviation d_r and it will be assumed that the base-band signal will be adjusted so that $d_a=d_r$. Also, at the present state of the art, it is usual to make full use of the baseband spectrum, so that $f_m\simeq Nb$, although N on this basis may in fact be slightly greater than the actual number of used voice channels.

With this understanding and using equation 2 we obtain

$$\psi_n=\frac{F_0r_0b^3XAN^{(a+2)}}{d_r^2W_0(8\cdot4)(10^{16})} \qquad (8)$$

If d_r is kept constant, irrespective of the number of channels which are to be used, then the way in which the weighted thermal signal/noise ratio in the top channel will vary with N is obtained from equation 8.

Nonlinear Noise and Noise Loading

The distribution of the noise which results from the transmission of a complex wave through a nonlinear device, and as particularly applied to single side-band suppressed carrier-type multichannel systems, was probably first worked out in detail by Bennett.[9] A more precise attack was made later by Jacobsen.[7] Subsequent authors have further investigated the problem.[8–4,10–13]

This work demonstrates that the total distortion power falling in the band in which signals are being transmitted is substantially independent of the powers of the individual signals comprising the complex multichannel wave. It is determined rather by the mean power of this composite signal. The distribution of the various noise components arising from the distortion is affected somewhat if the fundamental power is concentrated entirely in some portion of the frequency spectrum but in practice, the random selection of channels in an operative system mitigates against this. Consequently, the distribution is little different from that expected with a signal having a uniform power spectrum.

Thus a white noise signal is a convenient means of testing the system under dynamic conditions. The concept of using a white-noise signal in order to establish the level of intermodulation products was also suggested by Bennett. However, since that time the idea has gained considerable popularity and has been discussed by several authors.

The method has several advantages over the more conventional distortion measurements. First, the sum total of all nonlinear noise components is measured at once. Next, the test-signal waveform approaches that of the multichannel signal in an operative system. The test level can be such that the effect of overload due to energy peaks extending into the overload region, can be assessed at the same time. An additional feature is that it is possible to measure the unwanted components falling in any part of the band (with suitable filters), so that noise distributions can be determined with considerable accuracy.

The procedure is to use a white-noise signal with a noise bandwidth just equal to the base-band of the radio equipment. The noise which falls into a channel is then measured. This may be called the signal noise. A filter is next inserted which eliminates the energy transmitted in this channel so that the energy received is due only to intermodulation components (nonlinear noise). The ratio of the two then gives the signal/noise to nonlinear noise ratio in the channel. If a sufficiently strong r-f carrier is used, the influence of thermal noise can be eliminated and only nonlinear noise measured. Fig. 9 shows these noise bands.

Since the noise in the top channel is greatest, a system test actually only requires that a measurement be made in this channel. Further, because the signal falling in a specific bandwidth is definitely related to the total noise bandwidth used for the test, it is not absolutely necessary to measure the signal energy. This simplifies the measuring equipment to some

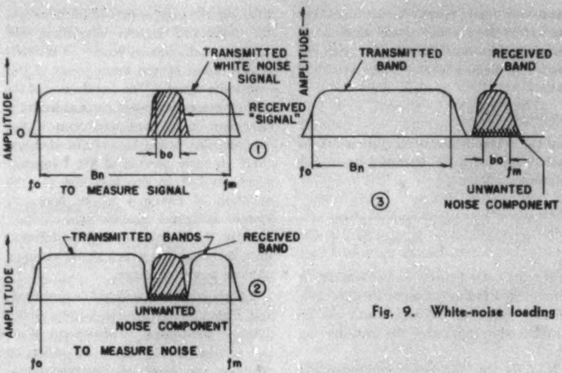


Fig. 9. White-noise loading

First method: The channel signal is determined from the band b_0 as in 1. The noise is then measured in the same part of the spectrum as in 2, a band-elimination filter being used to prevent direct transmission into the measured channel

Second method: The channel signal is (b_0/B_n) times the power in the band B_n. The noise is then measured in the band b_0 just above B_n as in 3

extent because a band elimination filter is no longer necessary. This method is also shown in Fig. 9. Of course, if it is required to establish definite figures for the channels within the band, then the former method is required. Because of its relative simplicity, the latter method is recommended for a system test.

Although the peak distribution is similar to that of a multichannel signal and the effect of overload can therefore be more readily examined, it will be found occasionally in this test that the overload point is difficult to define. It may then be necessary to relate the mean noise power to the overload point determined by means of a single-tone sine wave. This relation is given simply by P_p/P. Therefore it can be obtained by combining data in Figs. 5–8. The result of so doing is shown in Fig. 7. (Fig. 7 includes the allowance for limiting previously mentioned. If this is not required, then the curve for $\epsilon = 10^{-3}$ of Fig. 1 should be used.)

When a white-noise test signal is used, the effect is equivalent to having all channels in the system active, which, as has already been pointed out, does not necessarily apply to the operative system. Consequently, white-noise test data must be corrected to take this into account.

Let

B_n = power bandwidth of the white noise used to simulate the multichannel signal

γ_n = measured signal/noise ratio in a test band b_0, using white noise. Since the same instrument is used to measure both signal and noise in this test, this ratio is unweighted. γ_n is measured when the power in the band B_n is the same as the equivalent power of the composite multichannel signal; $b_0 \ll B_n$

Both the signal and noise spectra can be regarded as substantially uniform over small bands within B_n. Consequently, if b_0 is not too much different from the expected channel spacing b, we can expect that γ_n is the value which would also have been obtained if the test band b_0 had been equal instead to the channel spacing b.

Now the number of channels which could be fitted into the band B_n is

$$N = B_n/b$$

But the ratio between the equivalent power of the composite signal and the single-channel test tone is obtained from equation 2 so that,

$$P = XA(B_n/b)^{\alpha} \tag{9}$$

In the white-noise test, the ratio P_0 beween the mean power of the composite signal and that which would appear in the band b is,

$$P_0 = B_n/b \tag{10}$$

Hence, when the white-noise test signal power equals that expected from the composite signal, the channel signal power in the white-noise test is different from the single-channel test-tone power.

On the other hand, since the white-noise test signal has the same form and power as the composite multichannel signal, the nonlinear noise due to each, falling into the band b will be the same in each case. Thus, the test tone/nonlinear noise ratio S_c, expected in the operative system, in terms of γ_n obtained from a white-noise test signal, will be

$$S_c = \gamma_n P_0/P$$
$$= \frac{\gamma_n}{XA}\left(\frac{B_n}{b}\right)^{1-\alpha} \tag{11}$$

Because of its usefulness, equation 11 has been put in nomogram form in Fig. 10.

If, in the white-noise test the channel signal used is actually the band B_n then it is apparent that the value of signal/noise ratio measured would equal $\gamma_n \times (B_n/b_0)$.

The expected test tone/weighted nonlinear noise ratio S_c' is

$$S_c' = \frac{\gamma_n}{XAK}\left(\frac{B_n}{b}\right)^{1-\alpha} \tag{12}$$

In practice, the equipment may be used for a system of N channels while the noise test may be made for N_r channels. In order to establish the relation between the nonlinear noise which falls in the top channel of an N channel system in terms of N_r channel performance it is necessary to discuss briefly the distortion components in the FM system.

Distortion will arise due to both amplitude and phase nonlinearities in the various elements of the system. The former produces a noise spectrum which is essentially flat over the working bandwidth B_n. However, the latter produces an essentially triangular spectrum at the output of the discriminator similar to that obtained for thermal noise in the FM receiver. The resultant of these two noise distributions is a gradual transition from one to the other over the band, the net slope determined, of course, by the relative amplitudes of the separate effects.

It is inherent in the characteristics of these components that the total noise power from the separate distributions add power wise in each channel, although there may be voltage adding components associated with each one alone.

The distortion power at any frequency f in terms of that at the highest frequency can therefore be expressed by the relation $(f/f_m)^{\phi}$, where ϕ is obtained from the net slope referred to. Usually, over the range of frequencies in which interest lies, $\phi = 2$, and in general, $2 \geq \phi > 1$.

If the power for the test spectrum corresponding to the N channels is equal to that for the N_r channels, then the ratio of the nonlinear noise powers falling in the top channels of the systems will depend on the relative distortion powers which apply at the respective frequencies. Consequently, if γ_r is the signal/noise ratio measured in the top channel of the N_r channel system, then

$$\gamma_r=\gamma_n(f/f_m)^{\phi}$$

Again using the premise that the channels extend to zero frequency

$$\gamma_r=\gamma_n(N/N_r)^{\phi}$$

This may then be substituted in equation 12. Using ψ_c for the weighted noise to signal ratio, we then obtain

$$\psi_c=\frac{XAK}{\gamma_r}\left(\frac{b}{B_n}\right)^{1-\alpha}\left(\frac{N}{N_r}\right)^{\phi}$$

Since

$$B_n=Nb$$

$$\psi_c=\frac{XAKbN^{(\alpha-1)}}{\gamma_r b}\left(\frac{N}{N_r}\right)^{\phi}$$

As before, Kb is a constant $=1.5(10^{8})$ so the factor K may be easily eliminated, and we have

$$\psi_c=\frac{XA1.5(10^{8})N^{(\alpha-1)}}{\gamma_r b}\left(\frac{N}{N_r}\right)^{\phi} \qquad (13)$$

The Optimized Design

The total noise power which falls in the channel due to the radio link is the sum of the power from thermal and nonlinear sources. This can be expressed by the following relation

$$\psi_t=\psi_{n1}+\psi_{c1} \qquad (14)$$

Where ψ_t is the total weighted noise/test-tone signal (power) ratio, ψ_{n1} and ψ_{c1} are the thermal noise/signal ratio, and nonlinear noise/signal ratio respectively.

If no restriction is placed on the base-band signal transmission level, it may be readily shown that the minimum value of ψ_t is obtained when $\psi_{n1}/\psi_{c1}=\beta-1$ where β is the order of distortion producing the nonlinear noise.

In practice, the nonlinear noise is rarely due to contributions of a single order of distortion and the error if we let $\psi_{n1}=\psi_{c1}$ is very small. For optimum signal to noise ratio, therefore, it can be assumed that the nonlinear and thermal noise powers falling in the channel should be equal.

It should be remembered that both ψ_{n1}, ψ_{c1} are not fixed quantities but are determined by independent probability distributions. ψ_{n1} depends on fading and ψ_{c1} on traffic. However $\psi_t\neq\psi_{n1}'+\psi_{c1}'$ unless they are simultaneous probabilities. (Primes are used to indicate values with equal probabilities.) In the event of a long deep fade, then of course the maximum value of ψ_{n1} can occur during this time. However, it is generally true that the time of day during which such fades occur is a time when traffic is comparatively light and therefore ψ_{c1} is small. On the other hand, for short-term fading the probability of simultaneous maxima for $\psi_{c1}\psi_{n1}$ is extremely small. Consequently if the limiting values of $\psi_{c1}\psi_{n1}$ are such that $\psi_{n1}=\psi_{c1}$ the total noise under practical conditions will be given by

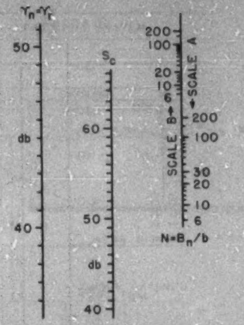


Fig. 10. Nomogram for equation 11. Scale A applies to integrated systems and scale B to nonintegrated

$$2\psi_{n1}>\psi_{ta}>\psi_{n1}$$

where ψ_{ta} is the total noise/signal ratio which can be expected under normal operative conditions.

It is reasonable to take advantage of this phenomenon in the design of the system and a compromise value for ψ_{ta} halfway between its limits will be taken as a reasonable one. Hence the optimum noise/signal ratio is given by

$$\psi_{ta}=0.75(\psi_{n1}+\psi_{c1})$$

If

$$\psi_{n1}=\psi_{c1}$$

Then

$$\psi_{ta}=1.5\psi_{n1} \qquad (15)$$

A system planned on this basis may be classed as an optimized system. Generally it is preferable to allot equal probabilities to ψ_{n1}, ψ_{c1} and this is recommended in fundamental planning.

For integrated systems, it would seem that the separate reliabilities should be not less than 99% so that the performance reliability would be 98%, using the busiest hour as base period. For nonintegrated systems it would seem that the separate reliabilities should be not less than 95% giving a resultant of 90%.

Now let the deviation d_c be the same as that which pertains when the ratio is measured. Next, increase the base-band signal power by the factor δ. If there is no restriction on the value of δ we can write for the new thermal-noise/signal ratio ψ_{n1} and the nonlinear noise/signal ratio ψ_{c1}

$$\psi_{n1}=\psi_n/\delta$$

$$\psi_{c1}=\psi_c\delta^{(\beta-1)} \qquad (16)$$

where β is the order of distortion producing the nonlinear noise. Because in fact the nonlinear noise may be due to more than one order of distortion, β does not have to be a whole number and in general $3>\beta>2$; β can be found by determining how ψ_c equation 13 varies with base-band signal power.

From equations 15 and 16, when

$$\psi_{c1}=\psi_{n1}$$

$$\psi_{ta}=1.5(\psi_n{}^{(\beta-1)}\psi_c)^{1/\beta} \qquad (17)$$

from equations 8, 13, and 17

$$\psi_{ta}=1.5\left[\left(\frac{F_0r_0b^2XAN^{\alpha+2}}{d_r{}^2W_0(8{\cdot}4)(10^{16})}\right)^{\beta-1}\times \frac{XAN^{\alpha-1}}{b\gamma_r}\left(\frac{N}{N_r}\right)^{\phi}\frac{(1{\cdot}5)(10^{8})}{}\right]^{1/\beta} \qquad (18)$$

In order to obtain a more even distribution of noise over the band in an FM system, use is often made of base-band equalization. With this, the high-frequency end of the band is emphasized in transmission and corrected in the receiver output. By this means, it is possible to improve the signal-to-noise ratio in the top channel. Let ζ represent the factor which gives the improvement which can be effected by such equalization. Then this can be included in equation 18.

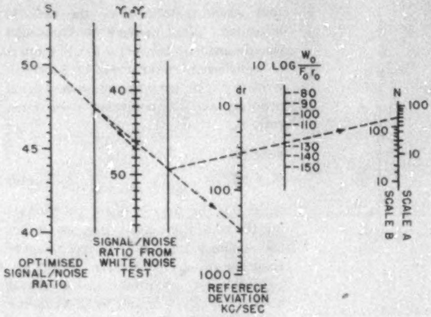
If we now simplify equation 18 and express the reciprocal as the signal/noise ratio S_t, we have

$$S_t=\frac{\zeta}{1{\cdot}5(X)(A)}\left\{\frac{d_r{}^2W_0(8{\cdot}4)(10^{16})}{F_0r_0}\right\}^{\beta-1/\beta}\times \left\{\frac{N_r{}^{\phi}\gamma_r}{(1{\cdot}5)(10^{8})b^{(2\beta-1)}N^{\beta(2+\alpha)-2+\phi}}\right\}^{1/\beta} \qquad (19)$$

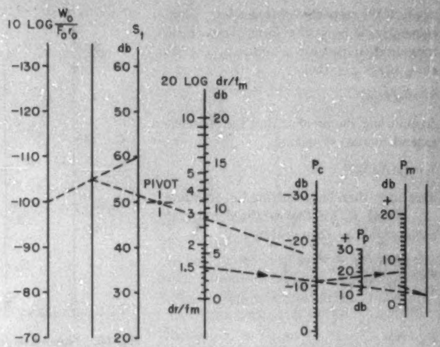
This now gives the optimum signal/weighted-noise ratio in the channel due to the radio equipment alone, in terms of actual system and equipment parameters.

If should be noted that equation 19 does not define the actual value of the final deviation corresponding to the composite signal. It is necessary therefore to establish that the peak power will not overload the equipment. This will be discussed later.

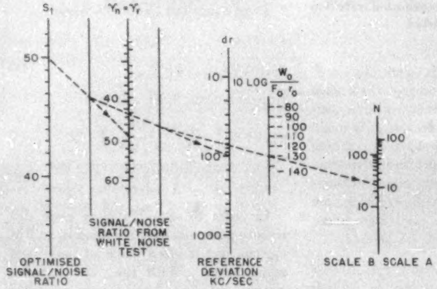
Although equation 19 seems rather formidable, substitution of numerical values greatly simplifies the equation. There are four principal cases of interest,



S_t
$\gamma_n = \gamma_r$
10 LOG $\frac{W_0}{F_0 r_0}$
dr
N
50
45
40
40
50
10
100
1000
-80
-90
-100
-110
-130
-140
-150
100
10
10
100
SCALE A
SCALE B
OPTIMISED SIGNAL/NOISE RATIO
SIGNAL/NOISE RATIO FROM WHITE NOISE TEST
REFERECE DEVIATION KC/SEC



10 LOG $\frac{W_0}{F_0 r_0}$
S_t
db
20 LOG dr/f_m
db
PIVOT
P_c
db
P_m
db
+
$+ P_p$
db
-130
-120
-110
-100
-90
-80
-70
60
50
40
30
20
60
50
10
5
4
3
2
1.5
20
15
10
5
0
dr/fm
-30
-20
-10
0
30
20
10
20
10
0



S_t
$\gamma_n = \gamma_r$
dr
10 LOG $\frac{W_0}{F_0 r_0}$
N
50
40
40
50
60
10
100
1000
-80
-90
-100
-110
-120
-130
-140
100
100
10
10
OPTIMISED SIGNAL/NOISE RATIO
SIGNAL/NOISE RATIO FROM WHITE NOISE TEST
REFERENCE DEVIATION KC/SEC
SCALE B SCALE A

Fig. 11 (above, left). Median carrier power (W_0/r_0) when nonlinear noise is due to second-order products. Scale A applies to integrated systems and scale B to nonintegrated systems

Fig. 12 (left). Nomogram similar to Fig. 11, but applicable when nonlinear noise is due to third-order products

Fig. 13 (above). Nomogram for equation 23. P_p is obtained from Fig. 8 so that P_m can be found. If this is greater than applies in the actual equipment then the system cannot be optimized but must be peak limited

in two groups. For each group $\zeta = A = 1$, $b = 4{,}000$ cycles, and $\beta = 2$ or $\beta = 3$. For the general case $\phi = 2$. The groups apply to the two traffic conditions given for integrated and nonintegrated systems. For the first, $\tau = 0.25$ $X = 0.2225$, $\alpha = 0.625$, and for the second, $\tau = 0.125$ $X = 0.035$, $\alpha = 0.77$.

Case 1. Traffic reliability 99%, activity factor $\tau = 0.25$ $X = 0.2225$, $\alpha = 0.625$.

$\beta = 2$

$$S_t = 3{\cdot}54(10^8)d_r\left\{\frac{W_0\gamma_r}{F_0 r_0}\right\}^{1/2}\frac{N_r^{2\cdot125}}{N} \qquad (20A)$$

$\beta = 3$

$$S_t = 1{\cdot}2(10^7)\left\{\frac{W_0 d_r^2}{F_0 r_0}\right\}^{2/3}(\gamma_r)^{1/3}\frac{N_r^{0\cdot66}}{N^{2\cdot3}} \qquad (20B)$$

Case 2. Traffic reliability 95% activity factor $\tau = 0.125$ $X = 0.035$, $\alpha = 0.77$.

$\beta = 2$

$$S_t = 2{\cdot}26(10^6)d_r\left\{\frac{W_0\gamma_r}{F_0 r_0}\right\}^{1/2}\frac{N_r}{N^{2\cdot27}} \qquad (21A)$$

$\beta = 3$

$$S_t = 7{\cdot}7(10^7)\left(\frac{W_0 d_r^2}{F_0 r_0}\right\}^{2/3}\gamma_r^{1/3}\frac{N_r^{0\cdot66}}{N^{2\cdot43}} \qquad (21B)$$

Equations 20 and 21 have been put in nomogram form in Figs. 11 and 12 for the case when $N_r = N$. The N/N_r correction is discussed later. One of the fundamental equipment limitations which must be recognized is the receiver threshold level. If W_0/r_0 is set to this in equations 20 and 21 the minimum signal/noise ration S_t for which the system should be designed may be obtained.

The Relative Channel Power P_c

Equation 19 gives the optimum performance in terms of measurements made at some reference deviation d_r, corresponding to some reference power in the composite signal. In order to obtain this optimum the reference signal was altered by the factor δ; but equation 19 does not specify exactly the final value of this signal, requiring only that the thermal and nonlinear noise powers be equal. In order to adjust the system, it is necessary to know the ratio between the final value of the single channel test tone power and the known reference.

Therefore let

P_c = ratio of the single-channel test-tone power to reference power (measured in the base band). The test-tone power is that which will lead to optimum performance as obtained from equation 19

It is apparent that, irrespective of P_c, equation 2 always applies, consequently

$$P_c = (P)^{-1}\delta$$

From 16, we have

$$\delta = (\psi_n/\psi_c)^{1/\beta} = 1.5\psi_n/\psi_{ta} \qquad (22)$$

which leads to two useful forms for P_c.

From equations 22 and 13

$$P_c = \left\{\frac{F_0 r_0 b^2 N_r^{\phi}\gamma_r N^{\beta-\phi}}{W_0 d_r^2(12{\cdot}7)(10^{19})}\right\}^{1/\beta}\frac{1}{XAN^{\alpha}} \qquad (23A)$$

From equations 22, 7, and 19

$$P_c = \frac{F_0 r_0 f_m^2 S_t}{\zeta d_r^2 W_0 5{\cdot}6(10^{16})} \qquad (23B)$$

In equation 23 S_t is the value obtained from equation 19. Equation 23 is especially useful and has been presented in nomogram form in Fig. 13.

The Peak Limitation

In equation 16 and subsequently in equation 19 it was assumed that there was no restriction on the value of δ. Under certain conditions therefore it is possible that the adjustment bringing thermal and nonlinear noise powers into equality might be such that the peak power could

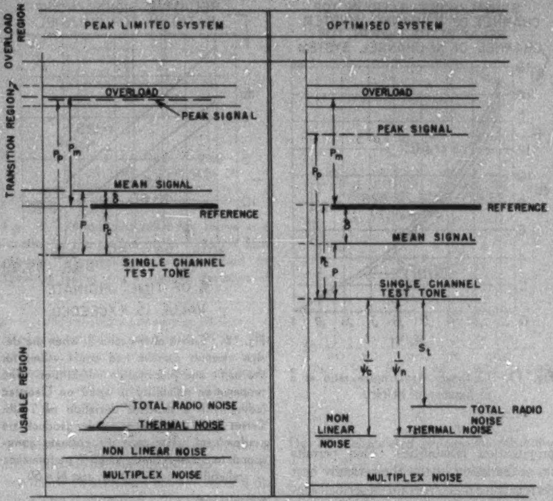


Fig. 14. The relative powers of the base-band signals in the multichannel system. Drawing not to scale

be considerably in excess of that permitted by the overload point of the equipment. Such a case could arise for instance, if the signal/noise ratio called for was rather small but the equipment produced little distortion. The test tone signal power obtained from 23 would then be unsuitable. The load capacity factor would have to be used and equation 3 or Fig. 8 would apply. However, it should be appreciated that the indiscriminate use of the load-capacity factor to so determine the single-channel-test tone power can reduce the maximum signal/noise ratio of which the system is capable.

Systems operated in accordance with this peak limitation may well be called peak-limited systems as distinct from the optimized systems under present discussion. Generally for the peak-limited mode, only thermal noise is significant and the performance reliability is almost determined by propagation. In this case the signal/noise ratio can be obtained in a form similar to equation 8 and is[6]

$$S_n=\frac{d_p^2(8\cdot 4)(10^{18})}{ZAN^{\alpha+2}b^2}\left(\frac{W_0}{F_0r_0}\right)$$

where d_p is the peak deviation corresponding to the overload point. Base-band signal levels and the various power ratios associated with the two modes of operation are shown in Fig. 14.

It is possible to establish an important system limitation as a function of the overload point. Let

P_m = ratio between the power at the overload point and the reference power corresponding to the reference deviation. P_m has a specific value for a given piece of equipment.

The ratio between peak power and single-channel test-tone is P_p. Also the ratio between test-tone power and reference power is P_c, hence the basic requirement is that

$$P_m \geq P_cP_p$$

P_c is given by equation 23 and P_p by 3.

When $P_m<P_cP_p$, the system cannot be optimized and the peak-limited mode applies. When $P_m=P_cP_p$ the system is both peak-limited and optimized, generally a rare occurrence and finally, if $P_m>P_c,P_p$ the system can be optimized. This requirement for P_m has been added to Fig. 13.

The General Relations With Optimized Systems

It is useful to extract some of the relations indicated by equation 19 to show the manner in which the signal/noise ratio is affected by various parameters. Since these relations can be obtained by quite simple algebraic manipulations only the end results will be shown here.

IMPROVEMENT IN SIGNAL/NOISE RATIO OBTAINED BY EQUATING THERMAL AND NONLINEAR NOISE

The total noise can always be obtained from $\psi_{nl}+\psi_{cl}$.

Using equations 14 and 16, the improvement ratio is

$$\frac{\delta^\beta+1}{2\delta^{\beta-1}} \tag{24}$$

where $\delta=(\psi_n/\psi_t)^{1/\beta}$ and ψ_n, ψ_c are the noise/signal ratios pertaining before equality is obtained.

EFFECT OF CHANGING THE MEDIAN CARRIER POWER (W_0)

If, when the carrier is altered, the deviation is likewise changed so that optimization always results, then equation 19 holds and

$$S_t \propto (W_0)^{\beta-1/\beta} \tag{25}$$

From equation 8 we can put

$$S_r \propto W_0 \tag{26}$$

which applies for the peak limited system.

Now suppose the system is optimized and the carrier altered without changing deviation so that optimization is not maintained. Then 25 can be used in 8 and 14 and we obtain an improvement ratio for S_t of

$$2k'/(1+k')$$

where k' is the ratio of carrier powers. These equations have been shown in Fig. 15.

It is perhaps useful to emphasize that if a carrier increase gives a proportionate increase in signal/noise ratio, then equation 26 applies. This is a sure indication that the system was not optimized in the first place and in all probability, the increase in signal/noise ratio could have been obtained without increasing the carrier at all.

Optimum signal/noise ratio as a function of reference mean-speech power can be derived from equation 19

$$S_t \propto 1/A$$

so that the ratio is inversely proportional to the mean-speech power.

OPTIMUM SIGNAL/NOISE RATIO AS A FUNCTION OF CHANNEL SPACING

From equation 19 S_t can be obtained in terms of N, N_r and b.

We can put $N=f/b$ and $N_r=f_m/b$. For a given system we can regard f and f_m as constant. This leads to

$$S_t \propto b^\alpha$$

This equation is illustrated by Fig. 16.

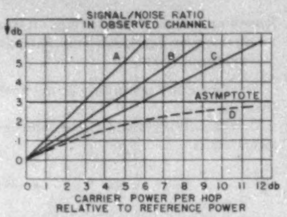


Fig. 15. Signal/noise ratio versus median carrier power

A—Noise due to thermal effect only
B—Nonlinear noise due only to third-order components
C—Nonlinear noise due only to second-order components
D—Improvement obtained if the deviation is not altered as the carrier power is changed. Curves B and C pertain to an optimized system

OPTIMUM SIGNAL/NOISE RATIO AS A FUNCTION OF THE PARTIALLY LOADED SYSTEM

Equation 19 gives S_t for the top channel of an N channel system, in terms of test data obtained for the equivalent of N_r channels. It is useful to observe how the expected top channel optimized signal/noise ratio varies with the value of N.

From equation 19 we get

$$S_t \propto \left(\frac{1}{N}\right)^{\alpha+2+\frac{\phi-2}{\beta}}$$

This is shown graphically in Fig. 17.

Performance Reliability

There are two aspects of performance reliability. First, it is possible to observe how the design value S_t varies when specific values are assigned to traffic and propagation reliabilities. This permits an examination of the theoretically best performance under certain prescribed conditions and it is on studies of this sort that basic economic concepts can be established.

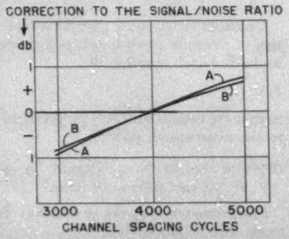


Fig. 16. The effect of channel spacing in the optimized system

A—Applies to nonintegrated systems
B—Applies to nonintegrated systems. The product Nb is constant

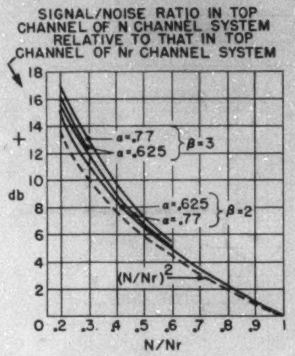


Fig. 17. Optimum signal/noise ratio as a function of N/Nr

Second, once the design is based on these specific values of the reliabilities, it is useful to observe how the signal/noise ratio might be expected to vary under actual operating conditions; that is, when the actual values of the reliabilities depart from the theoretical figures.

THE DESIGN VALUE OF S_t AS A FUNCTION OF SPECIFIC VALUES OF RELIABILITIES

From equation 19 we have the following relation

$$\underset{(N=N_r)}{S_t} \propto \overline{X}\left(\frac{W_0}{r_0}\right)^{\beta-1/\beta}\left\{\frac{}{N^{\alpha\beta+2\beta-2}}\right\}^{1/\beta} \tag{27}$$

Changes in S_t can be determined by assigning values of X and α (traffic reliability) and r_0 (propagation reliability).

This has been done to obtain the curves of Fig. 18. Each design value of S_t was calculated by making the propagation reliability equal to the traffic reliability. Also since second-order distortion is usually predominant, the curves are based on $\beta=2$.

Using the curve for $N=60$, $\tau=0.25$ for instance, we see that a design based on 99% performance reliability will give a value for S_t which is 6.5 db lower than one based on a performance reliability of 90%. It is evident that it might well be preferable to sacrifice the reliability than face the economics of obtaining the 6.5 db.

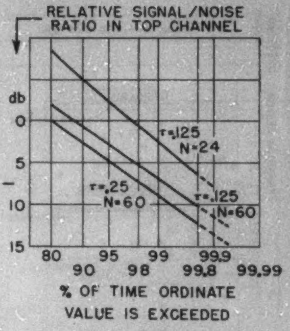


Fig. 18. Curves of the ratio S_t when the design assumes specific and equal values for the traffic and propagation reliabilities. The propagation reliability is based on Gaussian fading with a standard deviation of 7 db. Curves apply when second-order products are predominant. The zero db ordinate corresponds to a design based on 80% performance reliability when $\tau=0.25$, and $N=60$

INCREASING THE CARRIER POWER

Suppose that the propagation reliability is increased by increasing the carrier power, but the traffic reliability is unchanged. From equation 27, when W_0 is uncreased r_0 may be increased a like amount and the value of S_t will remain unchanged. However a larger value of r_0 is associated with a greater propagation reliability and this in turn affects the performance reliability. This effect has been shown in the representative curves of Fig. 19. It can be seen that increase in performance reliability by carrier power alone can be a depressingly unprofitable procedure.

EXPECTED PERFORMANCE RELIABILITY OF THE OPERATIVE SYSTEM

In the operative system we can use equation 8 to obtain the relation

$$\psi_n \propto r_0$$

The nonlinear noise/signal ratio, on the other hand, is dependent on the equivalent power of the complex signal, since this determines the nonlinear noise power. Consequently we can write

$$\psi_c \propto P^\beta$$

and then for the total noise/signal ratio ψ_t we obtain

$$\psi_t = k_1 r_0 + k_2 P^\beta \tag{28}$$

where k_1 and k_2 are constants. Hence the distribution of ψ_t is obtainable from the separate distributions of r_0 and P^β.

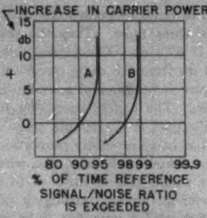


Fig. 19. Performance reliability shown as a function of the median carrier power at each receiver. Propagation reliability based on Gaussian fading with a standard deviation of 6 db

A—Traffic reliability constant at 95%
B—Traffic reliability constant at 99%. The zero db ordinate corresponds to that carrier power for which the propagation and traffic reliabilities are equal

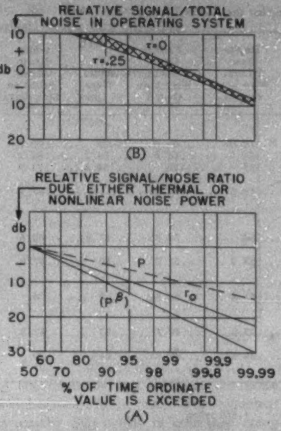


Fig. 20. Speculated performance reliability

A—Assumed distributions for r_0 and P
B—Resultant distribution using the curves of Fig. 20(A). The zero decibel ordinate corresponds to the calculated signal/noise ratio for the optimized system in which the traffic and propagation reliabilities are each taken to be 99%

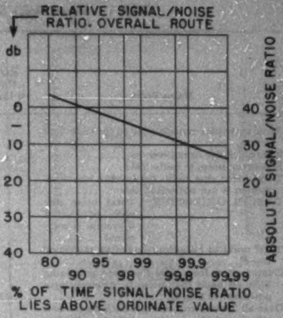


Fig. 21. Performance reliability of eight circuits in tandem, each with a distribution lying within the shaded area of Fig. 20(A)

The curves of Fig. 20(A) give suggested distributions for the busiest hour. These are based on standard deviations of 6 db for r_0 and 4 db for P. (The distribution shown for P includes the effect of signalling tones, carrier leak, and the like. It is apparent that certain types of signalling, for instance those in which a tone is present whether the channel is busy or not, may have a distribution with much smaller deviation, although the absolute value of P may be greater.) It is apparent that the percentage of total time the ordinate value is not exceeded is obtained by multiplying the corresponding abscissae values by 1/24. If the fading exhibits long term variations, the median value of W_0 (necessary to calculate the absolute value of the signal/noise ratio S_t from equation 19) should correspond to the median value pertaining during the hour of deepest fading.

The separate distributions for r_0 and P can be regarded as completely independent. If p_1 and p_2 are the separate probabilities that a given noise/signal ratio will be exceeded, then the resultant probability that this ratio will be exceeded will be $1-(1-p_1)(1-p_2)$. (It is important perhaps to note that this resultant is not the same as a Gaussian distribution with variance equal to the sum of the variances of the separate distributions, because the correlation coefficient is not zero.) The curve determined on this basis is shown in Fig. 20(B). The reference value of S_t indicated on this graph has been obtained from equation 15, equating the thermal and nonlinear noise powers pertaining to 99% reliability in each case.

Since deepest fading may occur when there is virtually no traffic, corresponding to $P=0$ in equation 28, the distribution of r_0 alone has also been indicated in Fig. 20(B). It can be expected that the performance reliability for the worst hour of any day will lie within the shaded area.

Basic System Requirements

It is possible to obtain basic system requirements with the aid of the previous data. In order to simplify the presentation of this information it is necessary to assign specific values to the factors in equation 19 wherever possible.

Accordingly, the data will be based on a fully loaded system, i.e., $N=N_r$. Also, the channel spacing will be standardized at 4,000 cycles. For the most advantageous design it is desirable to use some form of baseband equalization. A typical improvement figure for the top channel signal/noise ratio is 2 db so that ζ in equation 19 is 1.6. This factor is included in the recommendation below.

Signal/Noise Requirements for Integrated Systems

In order to establish the nonlinear noise power it is necessary to assume a certain traffic behavior. It would seem, in view of the early experimental work involved and its relation to actual systems, that the equivalent power should be that given by curve b Fig. 5. As seen from Fig. 19 there seems little point in designing for a propagation reliability greater than 99% so that the performance reliability aimed at should be 98%. This applies, of course, to the busiest hour period.

With respect to signal-noise ratio it can be expected that toll quality standards would be required. Some years ago the CCIF[2] tentatively suggested a fictitious circuit of reference as a guide in the design of long line circuits. This original model has been used in subsequent proposals for radio links and will be further discussed here. The basic reference circuit is a route of 1,550 miles in length comprising 2 audio patches. The over-all signaling noise ratio suggested is about 50 db, of which 25% is attributed to the actual channeling (multiplex) equipment.

A typical form of the expected signal/noise distribution is shown in Fig. 20(B). The noise contribution from the multiplex equipment has been neglected in this figure. However, since this noise is substantially less than the radio noise, it modifies the distribution only slightly and its effect can be disregarded in the ensuing discussion.

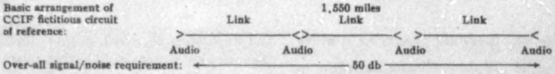
The longest conceivable route which could be used would be halfway round the earth, requiring eight of these CCIF links. These may be considered to have independent noise distributions lying within the shaded region of Fig. 20(B). The resultant distribution is shown in Fig. 21. The reference signal/noise ratio used is (50-10 log 8) = 41 db.

The absolute value of satisfactory signal/noise has not been adequately de-

Table I. Signal Noise Ratios for the Radio System

Noise From	Radio Over 0 Links	Radio Over 1 Link, Wire Line Over 2	Radio Over 2 Links, Wire Line Over 1	Radio Over 3 Links	1-Hop Radio Circuit Based on Link Data
Total multiplex contribution in the radio link above		60.8	57.8	56	60.8
Thermal noise in radio links		54.7	54.5	54.2	55.2
Nonlinear noise in radio links		59	56	54.2	71.2
Total noise due radio equipment above, if system is not optimized (b+c)		53.4	52.25		55.1
DESIGN OBJECTIVE: Total noise due radio equipment if system is optimized		54	52.35	51.2	60.8
Total noise of radio link including multiplex (a+e)		53.2	51.2	50	57.8
Contribution from wire line over remaining links on basis that noise is inversely proportional to length	50	51.8	54.8	0	0
Over-all route, considering both radio and wire line links, as measured at audio frequency	50	49.5	49.7	50	57.8
Propagation reliability required for radio links above, per cent		99.66	99.33	99%	99
Performance reliability, per cent				98%	98

Note: All ratios expressed in decibels, weighted.


Basic arrangement of CCIF fictitious circuit of reference:
1,550 miles
Link
Link
Link
Audio
Audio
Audio
Audio
Over-all signal/noise requirement:
50 db

fined, but a value better than 30 db in the worst channel for 99.9% of the busiest hour should certainly be acceptable.

On this basis it would seem that the CCIF requirement is rather greater than necessary. It is reasonable to expect that compandors would be used on a circuit such as that considered, and these could be expected to provide at least 15 db improvement. Allowing a substantial margin of error, it would appear that the absolute figure of 50 db for an individual link is some 10 db too great. (A more realistic objective, using Fig. 20(A) as a guide, would be a design value for S_t of 40 db with traffic and propagation reliabilities of 99%.)

When it comes to a single CCIF circuit, and again allowing a margin of error, it would appear that the CCIF figure of 50 db might again be reduced to 40 db without in any way sacrificing realistic performance. (More recent suggestions by the CCIF have in fact modified this figure.) The foregoing discussion has been purely on a speculative basis, so that the figure of 50 db will be retained in the subsequent discussions.

The noise in the CCIF circuit is generally taken to be inversely proportional to length. This is not a satisfactory assumption in the radio link. The design ordinarily uses a fairly large fading margin, so that the thermal noise at the end of the route is substantially independent of the number of hops. Thus, reducing the route length will make little difference to the thermal noise.

In order to put this on a quantitative basis, some assumptions about the fading must be made. Accordingly, the fictitious reference circuit will be supposed to consist of 50 hops, and the fading margin will be 20 db. The effect of the $H-1$ hops on the over-all thermal noise will be taken as equivalent to $(H-1)/2$ hops at median carrier level.

First consider that radio is used over the complete route. Since 25% of the noise can be in the channelling equipment, the S_t requirement for the radio alone will be 51.2 db. With equal thermal and nonlinear noise contributions the S_t ratio for either is 54.2 db.

Now consider that the radio equipment is to be used over only one link of the CCIF reference circuit, and the other two links will be wire line circuits, also conforming to the CCIF. The nonlinear noise in the radio link may be regarded as being inversely proportional to the length so that the value expected for this would be 54.2 +4.8=59 db. However, the number of hops has decreased from 50 to 17, and when the fading margin is taken into account, the expected signal to thermal noise would be about 54.7 db. Because these two ratios are not the same the system may be optimized, and from equation 24 the improvement which can be obtained is 0.5 db. Thus the total radio noise would give a signal/noise ratio of ≃ 54 db. This, therefore, should be the design objective for the 1 link radio circuit conforming to the CCIF circuit.

The channelling equipment is permitted to have a signal/noise ratio of 56 db, for three links. Thus the per link value is 60.8 db. Combining this with the 54 db from the radio circuit gives a net figure for the complete radio link of 53.2 db. The contribution due to the wire line part of the circuit, may be regarded as being inversely proportional to length. Thus for 2 links, the wire line section would give a ratio of 51.8 db. The over-all circuit, consisting of one radio link and 2 wire line links, would then have a signal/noise ratio determined from the 53.2 and 51.8 db figures, i.e. 49.5 db. If the over-all propagation reliability is set at 99%, then the per link value required is 99.66%. All of this information, together with similar data for two radio links, has been shown in Table I.

It is noted that on this basis, the over-all route signal/noise ratio is slightly less when the radio equipment is used over 1 or 2 links only. However, the effect is very small, and from the point of standardizing the system, may be neglected. Table I also shows the limits which might be imposed on a one-hop radio circuit using these CCIF circuit suggestions. It is probable that a propagation reliability of 99% would be acceptable for such cases.

EQUIPMENT REQUIREMENTS FOR INTEGRATED SYSTEMS

The equipment standards are perhaps best expressed by one link requirements. Using $\zeta=1.6$ in equation 19 and $S_t=54$ db in equation 20(A) we obtain

$$\frac{W_0}{r_0F_0}=\frac{19\cdot 6(10^{-3})N_r^{2\cdot 25}}{d_r^{3}\gamma_r}\quad (N=N_t) \qquad (29)$$

This is shown as a nomogram in Fig. 22, and is applicable when second-order products are predominant. Equation 20(B) may be used in a similar fashion for third-order products. Equation 29 calls for 99.66% propagation reliability and 99% traffic reliability. γ_r of course is the value obtained by a measurement over the complete link, and includes such effects as feeder and multi-path distortions. It is worth noting that the value of W_0 from equation 29 for a given equipment is some 20 db greater than would seem to be really necessary, as discussed in the section on "Basic System Requirements."

EQUIPMENT REQUIREMENTS FOR NONINTEGRATED SYSTEMS

The performance standards of nonintegrated systems are very closely controlled by economic factors, and it is a general principle of design to produce the most satisfactory circuit consistent with capital cost. Because this cost is in-

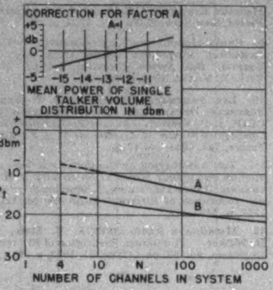
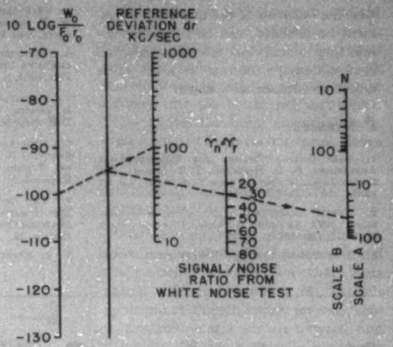


Fig. 22. Nomogram for equations 29 and 30 giving requirements for optimized systems

Fig. 23. Tone loading for system testing from equation 31

A—Curve for integrated systems

B—Curve for nonintegrated systems



timately bound up with the size of transmitters and the required antenna structures, the minimum median carrier level which can be used is of the utmost importance. This in turn is related to the estimated signal/noise ratio, as shown by equation 25. Consequently standards of performance in excess of real need and reasonable margin can be reckoned in terms of dead capital. In other words, the decibel per dollar can be established.

Curves similar to those of Fig. 20(B), but based on equating thermal and nonlinear noise powers at the 95% of the time values show that the reference signal/noise ratio required should be about 40 db, at which the performance reliability is 90% (during the busiest hour). (The minimum value for the signal/noise ratio is taken to be 30 db as discussed in the "Basic System Requirements" section.) Also the same proportion of noise as before will be attributed to the multiplex equipment so that the design value for S_t will be 41.2 db. Making the necessary substitutions in equation 21(A), including $\zeta = 1 \cdot 6$, we obtain

$$\frac{W_0}{r_0 F_0} = \frac{13.3(10^{-4}) N_r^{2(\alpha-1)}}{d_r^2 \gamma_r} \quad (N = N_r) \qquad (30)$$

when second-order distortion predominates. This equation is given as a nomogram in Fig. 22. Lower values of W_0 than this can sometimes be countenanced, and to assist in such determination equation 24 can be used. Equation 30 should be used in conjunction with 95% traffic and 95% propagation reliability.

System Tests With Tone Loading

One method of simulating the noise test is to inject a sine-wave test tone of constant frequency and amplitude into some of the channels and measure the noise which appears in the rest. The frequency is usually 800 cycles in countries using CCIF standards and 1,000 cycles in the United States.

If a small group of contiguous channels is energized, the nonlinear noise distribution may differ seriously from that to be expected with a uniform distribution of the signal energy and this effect will invalidate the measurements. For this reason, it is preferable to energize all channels except the one being measured, since the signal thus formed more closely approximates the uniform spectrum of white noise.

In order that the nonlinear noise power in this test should equal that expected from the traffic in the operative system, it is necessary that the mean power of the test signal should be equal to the equivalent power of the complex signal.

Let

P_t = power in the test tone applied to the N-1 channel inputs, relative to a 1-mw test tone

Then it is required that,

$$P_t(N-1) = XAN^{\alpha}$$

$$\therefore P_t = \frac{XAN^{\alpha}}{N-1} \qquad (31)$$

If one of the test tones is regarded as the signal, and the noise is measured in the top channel, then it is apparent that the measurement of the signal/noise ratio in this case corresponds to the factor γ_r in the "Nonlinear Noise and Noise Loading" section and the system performance can be predicted accordingly.

Satisfactory representation of the complex signal requires that the number of fixed tones should be fairly large, certainly greater than about 10. Results when a small number is used may therefore be in error. For this reason it may be desirable in some cases to inject more than one tone into each channel. Equation 31 is given graphically in Fig. 23.

Conclusions

It is possible to achieve an optimum design for the multichannel system by including factors related to traffic and propagation reliability. The performance reliability may then be shown as a function of the design values chosen for specific parameters. In addition, the reliability which might be expected in practice with the optimized system can be ascertained. This optimum condition is obtained by altering the per channel level to suit the given requirements and for a given equipment supplying a specific number of channels this will vary from system to system.

In order to establish the system design, it is necessary to define certain essential parameters. These include overload power and the distortion characteristics at some reference power which may or may not coincide with the former. With respect to distortion, it is necessary to establish the manner in which the total distortion varies with signal as well as determine the contributions from individual orders. This may be obtained conveniently by means of a white noise test signal and the basic capabilities of the system may be easily related to such measurements.

The actual performance requirements for systems which must become part of existing communication networks are necessarily more severe than those which do not. Standards are suggested which should provide a broad guide for the most economical planning of both these types.

An over-all system test, which requires

that the behavior of the multiplex equipment be included, can be conducted by injecting a sine-wave test signal into all channels except the one on which noise measurements are being made.

References

1. Load Rating Theory for Multichannel Amplifiers, B. D. Holbrook, J. T. Dixon. *Bell System Technical Journal*, New York, N. Y., vol. 18, 1939.

2. Comité Consultatif International Telephonique, XVI Assemblié Plénière, Florence, Italy, tome III, Oct. 1951.

3. Some Factors on the Engineering Design of VHF Multichannel Telephone Equipment, W. T. Brown. *Journal*, Institute of Radio Engineers, New York, N. Y., Feb. 1954, p. 51.

4. Intermodulation Noise on VHF Multichannel Telephone Systems, J. L. Slow. *Ibid.*, Feb. 1955, p. 67.

5. Interference in Multichannel Circuits, L. Lewin. *Wireless Engineer*, London, England, Dec. 1950, p. 294.

6. Non Linear Distortion in Transmission Systems, R. A. Brockbank, C. A. A. Wass. *Journal*, Institution of Electrical Engineers, London, England, pt. III, Mar. 1945, p. 1.

7. The Effect of Non Linear Distortion in Multichannel Amplifiers, B. B. Jacobsen. *Electrical Communication*, New York, N. Y., vol. 19, 1940, p. 29.

8. A Speech Volume Survey on Telephone Message Circuits, V. Subrizi. *Bell Laboratories Record*, New York, N. Y., Aug. 1953, p. 292.

9. Cross Modulation Requirements on Multichannel Amplifiers Below Overload, W. Bennett. *Bell System Technical Journal*, New York, N. Y., Oct. 1940, p. 587.

10. Les Faisceaux Hertziens a Courants Porteurs Devant les Recommandations du CCIF, J. P. Vasseur. *Annales de Radioélectricité Compagnies Francaises Associées de T. S. F.*, Paris, France, Jan. 1954, pp. 47–82.

11. The Application of Formulas for Distortion of Frequency Modulation to VHF Multichannel Radio-Telephony, J. Merriman. *Journal*, Institution of Electrical Engineers, pt. IIIA, 1947, p. 649.

12. Microwave Radio Links, A. T. Starr, T. H. Walker. *Proceedings*, Institution of Electrical Engineers, vol. 99, pt. III, no. 61, 1952, p. 241.

Silicon Rectifiers in Brushless Machinery

H. W. HENKELS
ASSOCIATE MEMBER AIEE

H. R. CAMP
NONMEMBER AIEE

J. R. FORTIER
NONMEMBER AIEE

W. J. SHILLING
ASSOCIATE MEMBER AIEE

R. E. SMITH
ASSOCIATE MEMBER AIEE

Synopsis: An illustration of the use of rotating silicon rectifiers in brushless machinery is given in application to generators for the *B-58* bomber. The evolution of both oil-cooled and air-cooled rotating rectifier assemblies is presented. The use of silicon rectifiers in this application illustrates their operation under the most extreme environmental conditions. In addition to all the usual salt spray, humidity, and shock requirements of the military specifications, the rectifiers were required to withstand centrifugal forces up to 15,000 *g* (acceleration of gravity) and oil pressures of 400 pounds per square inch, both at ambient temperatures from 150 to 180 C (degrees centigrade). Cells were required to withstand 175 volts in the reverse direction and 40 amperes in the forward direction. The brushless generators incorporating the rotating rectifiers were designed to deliver 40 and 60 kva (kilovolt-amperes).

THE idea of using rectifiers shaft-mounted on brushless machines has been considered for some time. Selenium rectifiers have been applied in some instances. In the aircraft industry, brush problems are becoming increasingly severe as planes fly at higher altitudes and greater speeds. Thus, in this industry, the need for brushless machines has been most urgent. The advent of the practical and reliable silicon rectifier has made possible the design of brushless machinery to operate under the most severe environmental conditions. This paper is concerned with describing the lines of rectifiers specifically developed for use in the rotating rectifier application. The basic silicon structures have been discussed.[1] The encapsulation of the basic units in suitable structures, the electrical and environmental testing of finished components, and the complete rectifier system in the alternator are discussed. Some information is given on the operation of the finished assembly. Further details on the generator developments will be found in references 2 and 3.

Two distinct alternator designs have been developed: an oil-cooled system, and an air-cooled system. Both have been completely tested, factory-produced in considerable numbers, and used without difficulty. The oil-cooled design has been operating with success in the *B-58* supersonic bomber.

Design 1: Oil-Cooled Rectifier

RING-MOUNTED DESIGN

Most large aircraft a-c generators have a rotating field with d-c excitation furnished by an integral d-c exciter, as shown in Fig. 1. To obtain a brushless machine with similar characteristics the best solution appeared to be to use a polyphase a-c exciter. The polyphase output from the armature could be rectified with cells integral in the rotor to provide direct current to excite the main a-c generator (see Fig. 2). The design of the main a-c generator could remain conventional and thus generator characteristics would not change appreciably. This solution meant, however, that a rectifier capable of high-speed rotation would have to be developed. A development program to obtain such a rotating rectifier was initiated.

The first rotating rectifier was made for a 40-kva oil-cooled (5-gallon-per-minute) aircraft a-c generator. The cooling oil at a temperature of 150 C passed through the shaft of the generator. It was required to design a rectifier cell and rectifier assembly which would be cooled by this oil. The first attack on the problem was that in which the six rectifier cells of the 3-phase system were bolted to an aluminum mounting flange. The cells were restrained by a 6-inch-diameter steel retaining ring to withstand centrifugal forces encountered at generator speeds up to 12,000 rpm. An analog of the system employed in evaluating temperature drops within the mount is shown in Fig. 3. The individual cells were required in the mounting described to withstand 15,000-*g* centrifugal force at the highest speed of the generator. The additional environment conditions will be described. Thus, the first basic silicon cell developed at Westinghouse was applied to the rectifier system required to operate at a cell temperature well over 150 C under the most extreme environmental conditions. The electrical, environmental, and size specifications for the unit are given in Tables I and II. In the following the design of the rectifier

---

Paper 57-793, recommended by the AIEE Metallic Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 27, 1957; made available for printing June 11, 1957.

H. W. Henkels, H. R. Camp and J. R. Fortier are with the Westinghouse Electric Corporation, Youngwood, Pa., and W. J. Shilling and R. E. Smith are with the Westinghouse Electric Corporation, Lima, Ohio.

The authors wish to acknowledge the contribution of persons at both Youngwood and Lima to the material presented; in particular, the work of S. J. Angello, H. Braun, W. Crapo, J. Crosby, and W. Green.

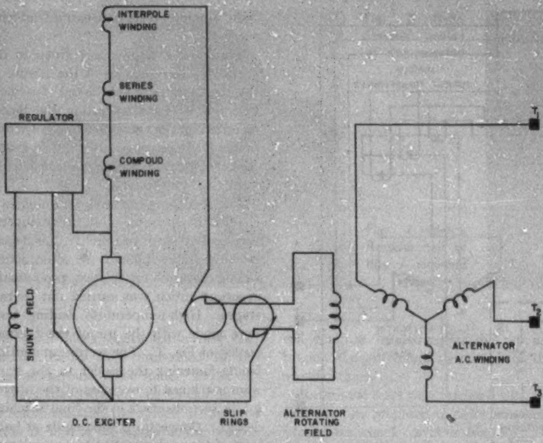


Fig. 1. Schematic diagram of conventional aircraft a-c generator

unit is considered first. Then are described the series of tests on the basic cell, the rectifier unit and the rectifier system in the generator to ensure satisfactory operation to the requirements of Tables I and II.

Rectifier Design

The basic cell was designed with 1/4-inch-diameter molybdenum counterelectrode contact, aluminum counterelectrode solder, *n*-type silicon over 10 ohm-centimeters in resistivity, silver alloy base solder, and 1/2-inch-diameter molybdenum base contact. Fig. 4 shows the basic cell in the lower right-hand corner. The structure and fabrication of this unit are completely described in reference 1.

The design of the rectifier cell or encapsulated unit began with the realization that the silicon should be in compression under the centrifugal forces. Fig. 4 shows, in the upper left-hand corner, the block that was developed for bolting to the mounting flange that was fastened to the shaft of the generator; see Fig. 3. A flexible member was required to connect the counterelectrode contact of the cell with the case terminal in order to prevent fracturing of the silicon because of mechanical and thermal stresses arising in temperature cycling and in operation. The flat flexible member shown next to the basic cell in Fig. 4 was designed to permit a free expansion in the direction perpendicular to the cell (and thus in the direction of the centrifugal forces) without allowing appreciable motion of the member in the plane of the counterelectrode. The flat flexible coupling thus provides:

1. The greatest possible motion of both the silicon and the external lead assembly with the least force (i.e., the spring constant is low), while at the same time providing a fairly rigid assembly to avoid cell stresses on starting and stopping.
2. A reasonable cross section for current flow in the flexible member.
3. A symmetrical member which will not tend to twist under centrifugal forces.
4. A connection which can be carefully cleaned of fluxes employed in soldering operations, thus ensuring reasonable life of the silicon cell.
5. A flexible member in the smallest possible space.

The seal chosen, shown next to the flexible coupling, was a "strain relief terminal" so that the glass would be in compression under the centrifugal forces and adequately supported by the metal flange of the cell. The case shown next to the stem assembly was designed to support the cell and stem properly in the smallest possible space. The flexible member was first soldered to the header with the use of flux, then very carefully cleaned of flux and baked out in vacuum. The cell contacts were pretinned with soft solder with the use of flux, then washed, and baked out. The surfaces were then mechanically abraded to brighten for the assembly operation. The stem assembly was joined to the basic cell without flux. This in turn was soldered, again without flux, to the pretinned cylindrical case in the same operation in which the header seal was made. The aluminum block was silver-plated and the cylindrical case soldered into the block with soft solder. Three of the aluminum mountng blocks were insulated from the mounting flange with mica.

Electrical and Environmental Testing

Static electrical tests of basic cells were made in several stages of processing and encapsulation to ascertain that units passed d-c specifications of 150-volt minimum at 18-milliampere reverse, 0.85-volt maximum at 1 ampere forward, and the units did not behave in an erratic manner. Rather complete forward and reverse static current-voltage curves were measured to obtain cells with reasonably regular characteristics. The electrical tests were frequently coupled with microscopic examinations of the silicon surfaces, particularly near the counterelectrode junction, for evidence of cracks of any kind. By development of suitable base and counterelectrode contact materials and solders, and with suitable processing conditions as described in reference 1, it was found possible to produce cells which were mechanically free of any visible defects and which were stable during the processing and encapsulation. In the course of development, the forward characteristics improved to the point where typical static values were: 0.75-volt drop at 1 ampere, and 0.85-volt drop at 10 amperes. Yields of units with suitable reverse voltages steadily improved.

Load and life tests were conducted on a

Table I. Rotating Rectifier Specifications, Design 1: Electrical Requirements

D-C Load				A-C Generator Load
Watts	Volts	Amperes	Period	Condition
1,100	55	20	continuous	full load
2,820	83	28	5 minutes	150% load
3,260	93	35	5 seconds	200% load
6,750	135	50	1 second	400% short circuit*
			Transient	
	175†	75†	0.04 second	removing 400% 3-phase short circuit
			0.02 second	applying 400% short circuit

* Power loss per cell: maximum of 20 watts in 400% short circuit.
† Peak cell voltage.

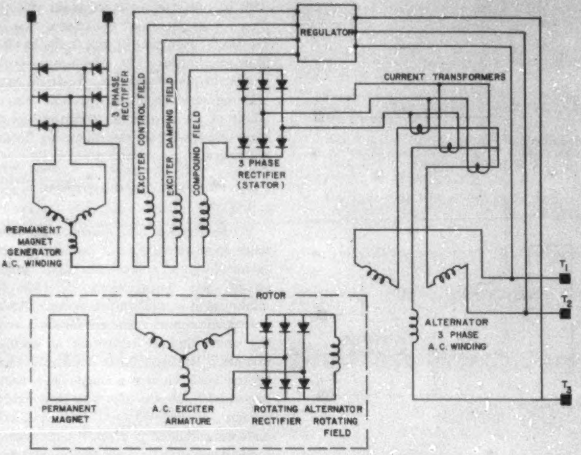


Fig. 2. Schematic diagram of brushless a-c generator

sampling basis. In Table III are shown 1,000-hour life test data on a set of six units in a 3-phase full-wave bridge. The test was performed with the following continuous operation parameters:

1. Case temperature, 180 C.
2. A-c applied voltage, 45 volts, 3-phase.
3. D-c load current, 20 amperes.

The units successfully passed the test specifications.

Temperature cycling tests were conducted on all cells and repeated on encapsulated units. The temperature cycling test incorporated temperature shock as follows:

1. Tested at 25 C.
2. Placed in 180 C oven on transite plate and tested when units reached temperature.
3. Tested at 25 C.
4. Placed in −55 C cold box on transite plate until units reached temperature.
5. Tested at 25 C.

This cycling test was also a shock test because the mass of the unit was so small that the rate of temperature change was extremely large. Initially the units were temperature cycled 20 times; however, subsequent statistical analysis indicated that 3 cycles were sufficient.

The forward current transient condition was not severe. The reverse transient voltage performance was checked by applying voltages through a capacitor having a time constant of over 0.02 second. Tests were made at voltages to 200 volts in an ambient of 180 C.

Immersion tests were conducted on all units produced to check against poor solder joints and cracked glass headers. Because of the inherent difficulties in soldering without the use of flux, 100% test was necessary. Flux was absent in all final assembly operations because of the desire to ensure good life test characteristics. The usual salt spray, altitude, humidity tests were carried out with success.

Particular attention was given to the vibration tests because of the nature of the application. The noted specification was followed in testing the block-mounted unit in three directions of vibration. No resonant points could be found. Neither electrical nor mechanical damage resulted from the tests.

The two most severe environmental conditions were coupled in high-temperature loading under rotation at 11,000 rpm. Because of the difficulty in constructing analog machines for the test, preliminary experimentation was carried out in two stages. High-temperature loading tests were made with the use of the thermal analog of Fig. 3. The rectifier mounting bands fastening the spider to the shaft were machined to sketches of the appropriate sections used in the final machine design. Connections were made at both ends of the analog section to a circulating pump and constant-temperature bath. An oil of properties similar to the engine cooling oil was circulated at rates in the range of those encountered in the engine design. The flow was determined to be in the laminar range. Since internal baffles were considered for the final machine design and the high-speed rotation would have caused considerable turbulance, the analog test gave conservative load ratings for the rectifier assembly. The complete analog was mounted in a temperature-controlled oven. Five of the cell blocks contained resistance heaters to supply up to the maximum permitted losses (20 watts per cell in 400% short circuit).

Table II. Rotating Rectifier Specifications, Design 1

	Paragraph Reference in MIL-E-5272A[4]	Procedure No.
Environmental		
Ambient operating temperature, −55 to 150 C Ambient pressures, 3.2 to 53 inches Hg High temperature	...4.1.1	...1: except 71 C to be 180 C; 50 hours to be 10 hours
Low temperature Temperature cycling, 200 cycles } −55 to 200 C	......4.2.1	...1
Temperature shock	........4.3.1	...1: except 85 C to be 180 C
Humidity test	........4.4.1	...1
Altitude test	........4.5.4	...4: except delete A; except B(1) −50 C to be 180 C; add B(2) 25 C ambient; except C(1) 9 hours to be 4 hours; add C(1) maximum altitude = 3.44 inches Hg, 180 C; add C(2) 25 C ambient
Salt spray	........4.6	...1
Vibration	........4.7.1	...1: except 160 Fahrenheit to be 180 C
Immersion	........4.12.1	...1
Shock	........4.15.2.1	...4: except 10 g in lieu of 15 g
Acceleration	........4.16.1	...1: 12,000 rpm for 1 minute. Mounted the same as in the application, acceleration to speed shall reach a maximum of 2,000 rpm per second
Life		
1,000 hours' peak inverse voltage of 55 volts; full load at sea level (dissipation doubles); total life 2 years.		
Size		
Basic unit without mounting block to fit in volume 1 by 1 by 1¼ inches.		

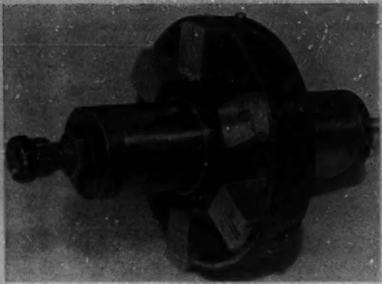


Fig. 3 (left). Thermal analog of ring-mounted system

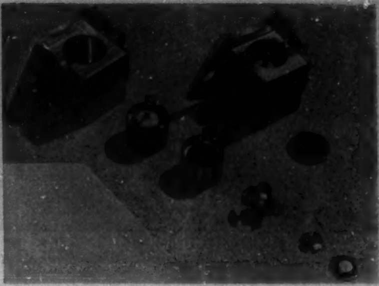


Fig. 4 (right). Rectifier cell of ring - mounted system

The sixth block contained a test cell. Thermocouples were embedded in the shaft at the base of the spider; in the rim of the spider between cell blocks, in the oil before reaching the spider, in the oven ambient, and in the solder at the bottom of the cell block holding the cell to the can. The test circuitry was arranged to measure peak and average leakages and forward drop.

The ambient air surrounding the cells was held at 180 C, and the circulating oil at 150 C. The tested cell was operated at peak voltages and at rms currents corresponding to those it would carry in the 3-phase connection. The load currents and reverse voltages were increased and the temperature drops in the system were measured. It was found possible to operate the cell continuously at the 200% overload rating (which was specified for only 5 seconds). Under this condition, the total temperature drop cell to cooling oil was 50 C. Of this total, a 21 C drop occurred in the oil film. Thus the shaft mount design and cell performance were more than adequate. Further, in final cell designs, the dissipation did not approach the maximum permitted in the specifications.

Rotational tests were initiated with the first samples prepared in order to get an immediate check on the feasibility of employing silicon cells in such high-speed rotating equipment at elevated ambients (180 C). Trouble was experienced with two or three of the first rectifier cells produced because of incompletely defined difficulties in both the testing apparatus and in the cells.

The remaining three cells of the first set of six (for the 3-phase, full-wave bridge) were tested with modified fixtures. Table IV gives data on reverse leakages and on forward voltage drops of these three cells connected in parallel. Leakages are at 100 volts, and forward drops at 30 amperes, both as functions of rotational speed and ambient temperature. The forward drop readings are not considered accurate because of brush difficulties. The reverse current readings are at the beginning and end of a 1-minute period.

After these preliminary tests, the cells were subjected to life tests under rotation. In these tests the ambient temperature was held at 180 C, the speed at 8,000 rpm. A reverse voltage of 75 volts d-c was applied continuously. Typically, the leakage of the cells read at the high temperature was satisfactory and reasonably constant or it decreased with time. Complete static characteristics were taken at room temperature at various periods during the course of the test. Fig. 5 presents the forward and reverse characteristics of a particular cell with parameters of temperature and periods of rotation at 8,000 rpm and 180 C. Tests were continued to 175 hours with success.

One of three cells life tested in the initial series showed an increase in forward drop at the end of the test. One of the most sensitive tests of a cell's condition is the forward drop at high currents. Units which exhibited such behavior were found usually to have defective soft-solder joints. All in all, the test results were very satisfactory. It was felt that 175 hours of life test were sufficient as a preliminary evaluation of cells under rota-

Table III. Life Test Data: Rotating Silicon Cell

Temperatures Given at Periods of Operation, Hours

Position	Ingot	Cell	180 C, 0 Hours	180 C, 100 Hours	180 C, 250 Hours	180 C, 350 Hours	165 C, 550 Hours	RT,* 650 Hours	177 C, 750 Hours	149 C, 850 Hours	RT, 0 Hours	RT, 1,000 Hours	180 C 1,000 Hours
Reverse Voltage† at 18-Milliampere Leakage Current, Volts													
1	*A-79-G*	*6A*	89	99	90	98	129	>150	100	107	>150	>150	78
2	*D-1-G*	*13B*	140	113	112	114	122	>150	115	118	>150	>150	76
3	*D-1-G*	*4*	120	100	95	100	106.5	142	96	100	>150	>150	102
4	*C-41-G*	*8C*	>150	>150	>150	>150	>150	>150	>150	>150	>150	>150	>150
5	*A-79-G*	*8E*	133	98	103	110	140	>150	118	122	>150	>150	124
6	*D-1-G*	*3*	>150	>150	>150	>150	>150	>150	>150	>150	>150	>150	>150
Forward Voltage Drop‡ at 10-Ampere Current, Volts													
1	*A-79-G*	*6A*	0.63	0.68	0.64	0.66	0.66	0.82	0.64	0.67	0.82	0.82	0.61
2	*D-1-G*	*13B*	0.79	0.82	0.81	0.84	1.04	0.86	1.17	1.04	0.86	0.98	1.00
3	*D-1-G*	*4*	1.06	0.73	0.70	0.73	0.70	0.94	0.77	0.77	0.94	1.00	0.89
4	*C-41-G*	*8C*	0.78	0.76	0.80	0.86	0.81	0.86	0.82	0.78	0.92	0.90	0.76
5	*A-79-G*	*8E*	0.83	0.84	0.84	0.83	0.88	0.91	0.92	0.97	0.91	0.95	
6	*D-1-G*	*3*	0.76	0.82	0.83	0.88	0.83	0.90	0.80	0.80	0.90	0.91	0.77

* Room temperature.

† For the reverse characteristics the 180 C operating temperature readings are the critical ones. All cells easily passed design 1 specifications throughout.

‡ For the forward characteristics, the RT readings are the most significant. Five cells passed the design 1 specifications before and after the test. The no. 3 unit indicated that it was poor and erratic at the start, yet this cell was easily below the allowed increase in output watts.

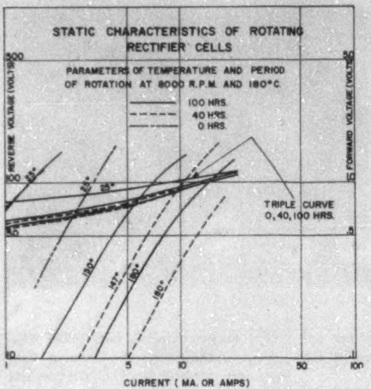


Fig. 5. Characteristics of ring-mounted cells on rotating life test

Table IV. Characteristics of Three Cells in Parallel Under Rotation

Temperature, C	Speed, RPM	Reverse Current at 100 Volts, Milliamperes	Forward Voltage at 30 Amperes, Volts
29	0	19/15	0.98
	6,500	18/16	0.96
	8,000	19/17.5	0.72
	10,000	26/22	0.62
100	0	13/11	0.98
	6,650	17/18	0.84
	8,000	22/24	0.82
	10,000	22.5/26	0.72
150	0	25.5/26.5	0.92
	6,900	36.2/44	0.87
	8,150	43/51.5	0.85
	10,000	44/61	0.67
180	0	80/89	0.9
	6,750	100/121	0.72
	8,125	80/104	0.78
	10,000	82/112	0.69

tion. The design objective was 1,000 hours' life (an aircraft generator requirement).

MACHINE OPERATION

These preliminary tests having indicated that the design was feasible, generators were built and tests made. The ring-mounted rectifier proved satisfactory in providing field excitation to a 40-kva 0.75-power-factor 208-volt 8,000 rpm aircraft a-c generator when the cooling oil temperature was below 130 C. The average output of the 3-phase bridge rectifier at this time was approximately 20 amperes. At oil temperatures higher than 130 C, the silicon cells short-circuited, causing loss of alternator load. The rotating rectifier had been designed to operate with 150 C cooling oil. Apparently, radiation from hot spots within the machine caused the heat conduction paths for the cells to carry more heat than was anticipated. This excess heat caused higher than expected temperature drops, resulting in severe cell temperatures.

To recheck the hypothesis that the cells were not failing at the originally specified ambient of 180 C, six special cells were produced which passed rectifier unit specifications at temperatures to 230 C. Even with the increased temperature range, machine operation was not successful with cooling oil at a temperature higher than 130 C.

SHAFT-MOUNTED DESIGN

Since in the application it was necessary to operate the generator with 150 C cooling oil, it was decided to change the mounting construction for the rectifier. A new construction was proposed in which the cells were mounted directly in the oil in the center of the shaft. In this mounting, the centrifugal force on the cells would be considerably lower than previously, and the cooling would be greatly improved. The mounting would be such that the cell would rotate about an axis perpendicular to the plane of the silicon wafer.

The principal changes in the rectifier cell specifications were those to require operation at 12,000 rpm for 5 minutes with the axis of rotation through the stem of the rectifier unit. The acceleration to speed was specified at a maximum of 2,000 rpm per second. The rectifier system in oil necessitated a rectifier unit test requirement of 50 hours' operation submerged in *MIL-L-7808* oil at 400 pounds per square inch pressure.

RECTIFIER DESIGN

The new specifications necessitated a reduction in physical size of the case and a change in lead arrangement, and a substitution of a compression glass seal. These changes were effected to produce a rectifier unit in a cylindrical case 1 inch high and 3/4 inch in diameter. The new mounting bracket developed is shown in Fig. 6.

ELECTRICAL AND ENVIRONMENTAL TESTING

It was realized that in the new shaft-mounted construction the silicon wafer might be off center. In this case centrifugal forces would cause shearing forces on the silicon wafer. Again, it was not known if these forces would cause cell failure. Preliminary tests were made to investigate this possibility. Three cells were placed in a fixture so that their cylindrical axes were off from the rotational axis by 0.010 inch. The three cells were rotated at 11,000 rpm for one hour and 14,000 rpm for one hour. Tests taken before and after rotation indicated no cell damage.

After the preliminary tests were complete, a rotating rectifier of the shaft-mounted design, Fig. 6, was made and assembled into a generator. This time, satisfactory operation was accomplished at full load, 40 kva with 150 C cooling oil. Overload and short-circuit requirements were also met with 150 C oil.

MACHINE OPERATION

At the time of writing, 25 oil-cooled machines have been manufactured and are operating satisfactorily with 150 C oil. Four machines have each successfully operated approximately 200 hours at various oil temperatures of from 38 C to 150 C in about 25-hour temperature cycles. One machine has operated in an altitude chamber at an ambient pressure equivalent to 52,000 feet for 8 hours and at an ambient pressure equivalent to 30,000 feet for 60 hours. These test results have been very satisfactory. Performance in a supersonic bomber has been excellent although, so far, experience has been limited to some 200 hours of generator flight time.

Design 2: Air-Cooled Rectifier

The development of rectifiers for the brushless oil-cooled generator gave much of the required experience to design a satisfactory rectifier for brushless air-cooled aircraft generators. The main difference in the application of the rectifiers is the manner in which they are cooled. The oil-cooled generator recti-

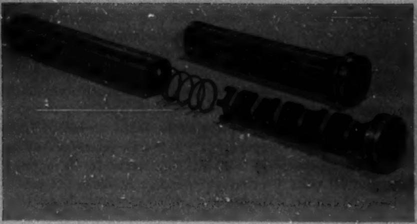


Fig. 6. Shaft-mounted oil-cooled assembly

Fig. 7. Rectifier cell of shaft-mounted air-cooled assembly

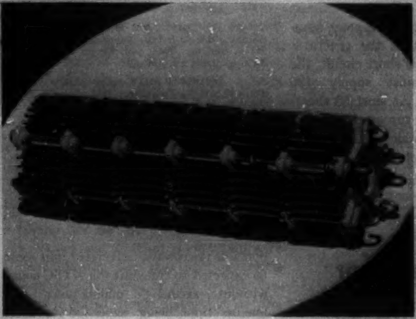


Fig. 8. Shaft-mounted air-cooled assembly

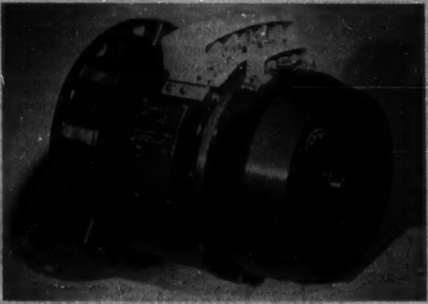


Fig. 9. Forty-kilovolt-ampere air-cooled generator

fiers are cooled just as well at high altitude as at sea level since the cooling oil flow and temperature are relatively constant. The cooling conditions for the air-cooled generator, however, vary greatly at different altitudes and aircraft speeds. The air-cooled generators must operate at altitudes from sea level to 65,000 feet with a cooling air temperature range of −55 to 120 C.

One satisfactory location for mounting the rectifiers had been determined from the oil-cooled generator design to be inside the shaft. This space is normally wasted on conventional generators and is dimensionally ideal for mounting the rectifiers. If blast-cooling air had not been available it would probably have been more desirable to ring mount the rectifiers, as this gives much better cooling. Since blast cooling was available, the shaft mounting was chosen because it results in less generator weight and length.

Table V. Rotating Rectifier Specifications, Design 2

	Watts	Volts	Amperes	Period
		Electrical		
D-c load	1,650	30	55	continuous
	2,450	35	70	5 minutes
	3,400	40	85	15 seconds
	5,500	100	160	10 seconds
Transient, peak cell		120	170	0.4 second 0.2 second
		Environmental		
Same as those of Table II				
		Life		
5,000 hours when bridge is supplying load of 30 volts and 55 amperes at case temperature of 150 C				

The electrical specifications for the rectifier system were altered somewhat to require lower cell-operating voltages and higher currents. The principal features of the revised specifications are given in Table V.

RECTIFIER DESIGN

To give an added factor of safety in the current-handling capabilities and to ensure minimum internal dissipation in the rectifier system (more critically important in the air-cooled design), the size of the basic cell counterelectrode was increased to 1/2-inch diameter. This change, which greatly increased current ratings, was accomplished by increasing the diameter of the case to 7/8 inch. Fig. 7 shows the parts of the air-cooled design. The stem and seal assemblies were changed to be consistent with the increased current ratings. As in the previous design, the assembly operation is carried out without the use of fluxes. The canned unit is pressed into the finned housing for mounting inside the generator in the air blast stream.

The shaft-mounted assembly for a brushless 40-kva generator is shown in Fig. 8. It is approximately 2 inches in diameter and 7 inches long. The fins are necessary to cool the rectifier under high-temperature operation during which the cooling air temperature may read 120 C.

ELECTRICAL AND ENVIRONMENTAL TESTING

In the course of continued development of the basic units, the static electrical characteristics have steadily improved. Typical forward drop values of the 1/2-inch cell are 0.8 volt at 30 amperes, and 1.22 volts at 300 amperes. Electrical and environmental tests of rectifier units have been revised and continued with success.

To determine if satisfactory cooling of the complete rectifier assembly were obtainable, tests were run at sea level with the rectifier mounted in the shaft and the shaft held stationary. Thermocouples were placed on the case of each cell. An external source of 3-phase a-c power was supplied to the rectifier assembly. The load was the main alternator field, which is the load in the application. These tests showed that the rectifier assembly could continuously supply the necessary field power (50 amperes at 30 volts) for rated generator output of 40 kva with 10 cubic feet per minute cooling air at 30 C flowing through the assembly. These results indicated that the system could be satisfactorily cooled under all the required altitude cooling conditions.

Machine Operation

Tests were run on the generator in the altitude chamber and the operation was satisfactory at sea level with 120 C cooling air and at 65,000 feet with −55 C cooling air. A 500-hour life test was then run with a cooling air temperature of 120 C and the operation at the finish of the test was almost identical with the operation at the start.

It is a requirement for aircraft generators to deliver 150% overload for 5 minutes and 200% load for 5 seconds over the altitude temperature range given previously. Under these generator load conditions the d-c field power required from the rectifier assembly is 70 amperes at 40 volts and 82 amperes at 50 volts respectively for the brushless 40-kva a-c generator. These overloads have been applied to one 40-kva unit over 50 times, thus far with no detrimental effects. There was an additional requirement for this generator that it deliver a minimum of 300% rated current on a 3-phase symmetrical short circuit at the generator terminals. During this short circuit, the rectifier assembly must supply 120 amperes at 72 volts. At least 50 short-circuit tests have been run on the generator without a single failure.

The results of the tests on the brushless air-cooled 40 kva were satisfactory to the point that designs immediately proceeded on other ratings. At present, brushless 30- and 60-kva generators are being tested and several other ratings are in the design stages.

A complete brushless, air-cooled 6,000-rpm aircraft 40-kva a-c generator is shown in Fig. 9. This generator is approximately 9$^{3}/_{4}$ inches in diameter and 12 inches long.

References

1. The Fused Silicon Rectifier, Herbert W. Henkels. *AIEE Transactions*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 733–46.

2. An Oil-Cooled A-C Generator for High-Speed High-Altitude Aircraft, H. J. Braun, W. J. Shilling. *Ibid.*, vol. 74, pt. II, 1955 (Jan. 1956 section), pp. 456–60.

3. A Brushless Air-Cooled Aircraft A-C Generator, R. E. Smith. *Ibid.*, vol. 76, pt. II, Sept. 1957, pp. 189–92.

4. Environmental Testing, Aeronautical and Associated Equipment. *Military specification MIL-E-5272A*, U. S. Bureau of Aeronautics, Washington, D. C., Nov. 16, 1953.

A Modified Microwave System for Color Television Transmission

F. F. McCLATCHIE
ASSOCIATE MEMBER AIEE

THE author's company has a continuing operational requirement for approximately 30 portable microwave radio relay systems in the Southern California area These relay systems are used to extend the Bell System broad-band network to temporary service locations in the area. to bridge existing facilities in the event of transmission failure, or to provide semipermanent studio to transmitter links.

These portable microwave radio relay systems may be briefly described as frequency modulated (FM), broad-band transmitters and receivers operable in the 6- to 7-kilomegacycle band at one-watt radio-frequency power output. The intermediate-frequency amplifiers are 20-megacycles (mc) wide and are centered at 130 mc. The receiver noise power is approximately −85 dbm (decibels referred to 1 milliwatt) which corresponds to a noise figure of 16 db.

The principal subscriber demand resulting in the use of these portable microwave radio relay systems has been for the transmission of monochrome television signals. While the transmission requirements specified for this signal are most exacting it has been possible to meet the demand on a routine basis. Our requirement is, given a service request, to expect that any system in stock can be assigned for service without excessive prior line-up time being expended.

With the advent of color television, and the adoption of the National Television Systems Committee (NTSC) standard color television signal, the transmission requirements became such that no existing portable microwave radio relay unit could be expected to transmit the NTSC signal without excessive maintenance and line-up time being expended to meet routine service requirements. The color quality of the NTSC signal is controlled by two transmission parameters which have not been markedly significant to the proper transmission of monochrome television pictures. These two parameters are specified as differential phase and differential gain.[1] In the particular case of NTSC color signal transmission these parameters specify what the variation of the phase and amplitude of a 3.579545-mc subcarrier is when it is superimposed on a larger amplitude low-frequency signal.

Anticipating that we would be faced with an increasing number of requests to carry NTSC color television signals, an investigation was initiated to determine the feasibility of improving our plant either by the purchase of new equipment or by the modification of existing equipment. Our purpose was to reduce the cost of maintenance by improving the plant to a point where NTSC color signal transmission demands could be met in the same routine manner as we had previously met the transmission requirements for monochrome television services.

Studies made on existing system types revealed that all system types available for investigation were deficient when considered with reference to the differential phase and differential gain aspects of the

---

Paper 57-918, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Pasco, Wash., August 28–30, 1957. Manuscript submitted May 8, 1957; made available for printing June 19, 1957.

F. F. McClatchie is with the Pacific Telephone and Telegraph Company, Los Angeles, Calif.

The author wishes to thank J. H. Clark and R. H. Coe for their contributions to the system redesign and for their aid and counsel throughout this project. The author also wishes to acknowledge the help and contributions of the members of the field and operating forces of the Plant Department, Southern California Area of the Pacific Telephone and Telegraph Company. Personnel of this department aided in the editing of the instruction manual and in providing some basic system objectives.

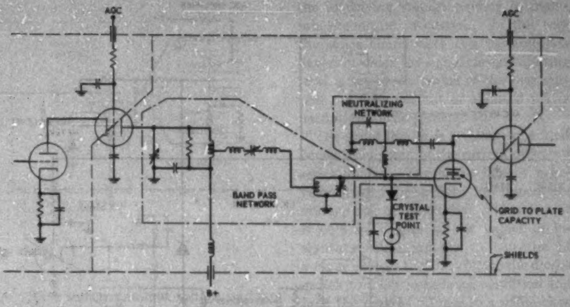


Fig. 1. Cascade i-f amplifier showing band-pass network, bridge T band rejection neutralizing circuit, and crystal test point

system transmission requirements. Various technical expedients were attempted to minimize these deficiencies. One of these expedients was the introduction of phase equalizers for the intermediate-frequency amplifier to produce a substantially constant delay through the center portion of the intermediate frequency pass band. Another expedient was to reduce the frequency deviation introduced by low-frequency components thereby decreasing over-all deviation without introducing excessive high-frequency noise. It was necessary in most cases investigated to simultaneously employ both techniques. The conclusion was that the expedients necessary to up-grade existing system types to meet our requirements were unsatisfactory, both from a transmission quality viewpoint as well as from the economic viewpoint of maintenance personnel time required to achieve satisfactory results.

The principal suppliers of equipment of the type being considered were questioned relative to their plans for producing a unit which would be satisfactory for our purposes. Encouraging responses from a technical viewpoint were received to the effect that the transmission requirements would be met by new systems in initial design stages but no precise information was available on the important factors of when the systems would be available, how much the units would cost, how the technical requirements would be met, or what the maintenance aspects of the systems would be. We therefore investigated upgrading our existing plant to meet our requirements. Preliminary investigation indicated that for an approximate investment amounting to one fourth of the initial costs of the relay systems we could improve our existing plant to meet all of our transmission, plant cost, maintenance cost, and availability time requirements. A detailed study was therefore initiated on our existing plant.

The modification of our existing systems was considered with respect to the following points:

1. System transmission characteristics without pre-emphasis or i-f (intermediate-frequency) phase equalizers
    - (a) The over-all system gain frequency characteristic to be essentially flat from 5 cycles to 8 mc.
    - (b) The over-all system differential gain to be less than 1 db.
    - (c) The over-all system differential phase to be less than 3 degrees.
    - (d) The receiver noise power to be less than −85 dbm. Equivalent to a noise figure of 16 db.
    - (e) The 60-cycle residual interference to be minimized.
    - (f) The system to be well shielded from r-f (radio-frequency) interference.
    - (g) The base-band noise spectrum to be essentially triangular and free from spurious tones.
2. System operational characteristics
    - (a) The receiver local oscillator to be controlled to seek a frequency within 1/2 mc of the frequency which produces the appropriate i-f resultant.
    - (b) There shall exist no built-in electrical or other hazards for the operator.
    - (c) A minimum of operator time to be expended on system equalization.
    - (d) A complete set of written operational instructions to be provided with the completed unit.
3. System maintenance characteristics
    - (a) All system maintenance alignments to be accessible without removing subcomponent covers.
    - (b) All major system subcomponents to be interchangeable by virtue of unitized plug in construction.
    - (c) Sufficient integral test points and detectors to permit simplified precision alignment of the system.
    - (d) A complete set of written maintenance instructions to be supplied with the completed unit.
    - (e) Due care to be given to the selection and placement of parts to minimize electrical hazards during maintenance operations.

System Transmission Characteristics

The microwave radio relay systems have been generally described in the introduction.[2] The transmitter unit was acceptable for NTSC color signal transmission but the receiver unit did not meet the differential phase and gain objectives. Reference sources and independent investigation indicated that the greater portion of the distortion was introduced in the i-f section of the receiver. Initial efforts were directed toward improvements in this area.

I-F Amplifier

The direct approach toward eliminating distortion in the i-f system would be to eliminate the i-f section itself and construct a tuned r-f receiver rather than a superheterodyne receiver. Suitable components such as traveling wave-tube amplifiers, wave-guide preselectors and discriminators are available to implement this solution. While this approach was feasible it was rejected because of arguments in favor of a certain solution possible using conventional techniques.

While no claim to originality is made with respect to the design reasoning used on this i-f amplifier it is felt that it is sufficiently unique to bear detailed explanation.

The basic requirements for the i-f amplifier were stated. These are:

1. Accept and pass a discrete set of frequencies (120 to 140 mc) with a minimum of phase, gain, or signal/noise degradation.
2. Reject all frequencies outside of this set.
3. Raise the level of the desired set to a point where the information content can be conveniently extracted, consideration being given to match the amplifier input and output to preceding and succeeding elements.

These requirements were transformed to the simplest physical system possible. This was a band-pass filter with inherent gain.

The interelectrode capacities of the vacuum tubes were utilized as filter section components. Investigation of the transfer characteristics of various simple filter sections indicated that a constant *k*, *pi* section, band-pass filter adjusted to

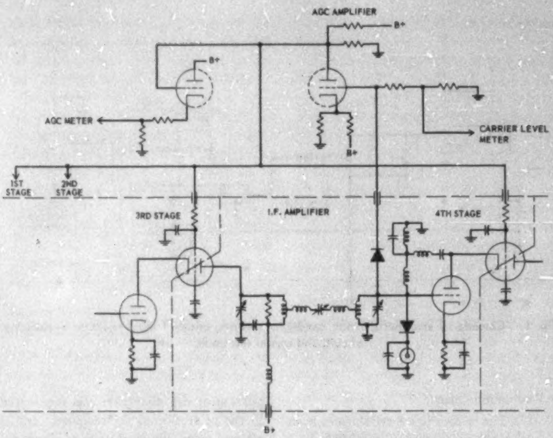


Fig. 2. Automatic gain control circuit

approach a Butterworth performance curve would be the most expeditious compromise between phase linearity, and adjacent channel rejection. An investigation was instituted to obtain the highest gain per stage that was consistent with stability, using available long-life tubes.

The Western Electric type *5842/417A* miniature triode was chosen because of its high transconductance (28,000 micromhos), low noise (gold-plated grid), and long life (10,000 hours). Experimental cascode amplifier stages were built which resulted in gains up to 24 db per stage at 20-mc band pass. In order to insure stability, the gain was reduced to 18 db per stage at 20-mc band pass.

Plate-to-grid feedback was reduced by incorporating the plate-to-grid capacitance of the vacuum tube into a band rejection, bridged *T*, filter network; see Fig. 1. This network improves the gain and stability figure over that possible with the more conventional parallel resonant neutralizing method. Simple resonant neutralization cannot be effective over the 20-mc pass-band range because, if the *Q* is high enough to isolate the grid from the plate, the band width of neutralization is too narrow.

High-gain per stage permitted the construction of an i-f amplifier with a total gain of 124 db requiring only seven band-pass interstage coupling networks and one transformer coupling network. High over-all gain insures that the limiter will be in full operation on noise alone after considerable aging of tubes. In practice, the transconductance of all the i-f tubes may be reduced 30% without impairing the limiter function. Tube life has been extended by operating the plate at 35% of the maximum rated dissipation and by providing heat conducting tube shields.

The interstage coupling network is shown in Fig. 1. The characteristic impedance at the input (plate) side of the coupling network is 1,500 ohms. The impedance of the series resonant coupling section is 75 ohms. The impedance at the output (grid) side of the network is 450 ohms. The input of the network is adjusted to 1,500 ohms to provide the highest gain, consistent with stability, that the tubes can deliver. In order that the series resonant elements be realizable values of inductance and capacitance, the impedance of this section must be kept low. The 75 ohm value was chosen as a compromise between ease of impedance matching and realizability of component values. The output side of the network was made as high as possible, consistent with transit time loading by the vacuum tube in order to sustain the least voltage loss through the network. The voltage reduction due to impedance transformation in the network is 5.2 db, other losses associated with the network are about 3 db, therefore the gain seen at the plate of the tube must be about 28 db to realize a stage gain of 20 db.

A desirable condition obtains with the iteration of identical (with compromises necessary for impedance transformation) networks. As various stages of the amplifier provide varying proportions of the over-all i-f gain under field operating conditions (path lengths vary from less than 1 to 100 miles) the various stage transfer characteristics assume different degrees of significance with reference to the over-all i-f amplifier transfer characteristic. If all stages are of like type, no change is apparent in the over-all transfer characteristic due to relative signal level. It was possible for the system to meet the following test because of this feature of the i-f amplifier.

The complete microwave system was aligned without external equalization to pass an NTSC color signal with 1/2 db differential gain and two degrees differential phase distortion. All i-f amplifier tubes were replaced with a random selection of new i-f amplifier tubes. There was no perceptable change in the NTSC color signal transmission characteristics after the tube replacement. This feature has not been characteristic of other systems. In fact certain systems must be returned to the factory for realignment after three tube replacements.

Automatic Gain Control (AGC)

The i-f amplifier must operate over a dynamic range of input levels of up to 50 db without incurring video level changes or excessive changes in the differential phase and gain parameters. Level changes of this magnitude would cause several i-f amplifier stages to overload on strong signals and cause changes in the transfer characteristic. These level changes cannot be compensated for with a manual i-f gain control because fading might occur at unexpected intervals. An AGC has been added to correct for these level changes. The gain of the i-f amplifier is set to maintain a constant level into the limiter circuit. This also serves to improve the action of the limiter circuit by restricting the dynamic range it must be able to handle. The AGC circuit derives its signal level indication from a germanium rectifier connected to the input of the fourth main i-f stage; see Fig. 2. This rectified positive voltage is filtered and applied to the grid of the AGC amplifier. This tube is biased to cut off by an augmented cathode bias and the plate voltage is derived from a voltage divider that allows 100-volt positive to appear at the plate when the tube is cut off. The plate of the AGC amplifier is connected to the grids of all the grounded-grid stages of the i-f amplifier through suitable r-f isolation filters. When the

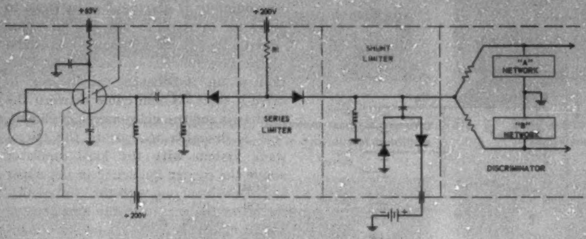


Fig. 3. Limiter circuit

signal voltage rises above a certain minimum value, the AGC control tube is caused to conduct in proportion to the signal intensity. This causes the plate voltage of the AGC tube and the grid voltage of the i-f amplifiers to become less positive and thereby reducing the gain of the amplifiers. In practice the voltage at the grids of the i-f amplifiers will vary from 100 volts at zero signal to about 40-volt at maximum signal.

Limiter

The function of a limiter in an i-f system used to receive FM signals is to establish a constant r-f level input to the discriminator regardless of carrier level or frequency. To function properly on NTSC color as well as voice carrier and pulse signals a limiter must meet these conditions:

1. Limit all frequencies within the i-f pass band to a single voltage level.
2. Suppress amplitude modulation present on the carrier at all video frequencies.
3. Present a very low driving-point impedance for the discriminator.
4. Have sufficient output voltage to drive the discriminator rectifiers for linear operation.

The usual vacuum tube limiters were deficient in one or more of the points noted above. The requirement of AM (amplitude-modulation) suppression at the color burst frequency (3.58 mc) was particularly severe and can be met only with circuits of inherently short time constant.

Biased diode limiters can meet the short time constant requirement. New developments in germanium diode technology made possible a diode-type limiter of superior characteristics. Diodes with a forward conductance of 300 milliampere (ma) at 1 volt (best available previously were 30 ma at 1 volt) were used as both series and shunt limiters. See Fig. 3 for schematic of the circuit used.

A forward bias current flows in the series limiter, allowing low instantaneous i-f currents to flow to the shunt limiter with little attenuation. When the instantaneous value of the i-f current exceeds the bias current, the impedance of the series diode rises proportionately with the i-f driving voltage tending to maintain a constant i-f current. In this way the series limiter acts as a variable loss device in which the signal current limiting level is set by the value of the bias current. A bias current of about 3 ma is used in this arrangement.

The shunt limiter is a voltage operated device as opposed to the current operated series limiter. The shunt limiter is reverse biased by a fixed low-impedance voltage source (batteries). This circuit is high impedance to low values of instantaneous i-f voltage. When the i-f voltage exceeds the bias voltage the diodes switch to a low-impedance condition and hold the peak i-f voltage to a value close to the bias voltage.

The shunt and series limiters operate together so when the i-f signal exceeds a predetermined value at the limiter input the series limiter switches to a high impedance state while the shunt limiter switches to a low impedance condition. This results in a very "hard" limiter action. Since the diodes switch on positive and negative portions of each i-f signal the time constant is easily low enough to suppress any video frequency AM. The limiter contains no frequency-sensitive elements that are selective within the i-f band pass, therefore all frequencies will be limited at the same voltage. The shunt limiter presents a low driving point impedance and delivers a 3-volt peak-to-peak i-f signal to the discriminator.

The noise power at the input of the receiver is −85 dbm when using the *IN23C* mixer crystal. Full limiting requires +18 dbm output from the i-f amplifier, therefore, a minimum gain of 103 db is required to obtain full limiting on noise alone. Since the over-all i-f amplifier gain is 124 db, a margin of 11 db

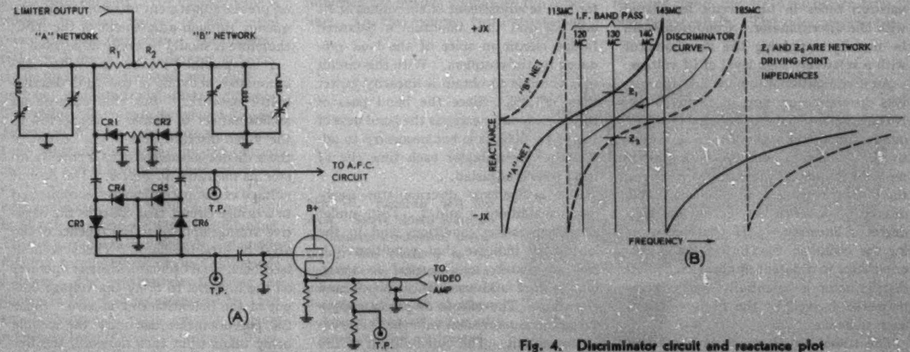


Fig. 4. Discriminator circuit and reactance plot

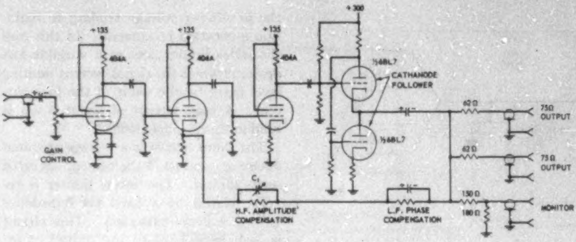


Fig. 5. Video amplifier output circuit

is established. This margin allows a 30% reduction of transconductance of all i-f tubes before the system fails to limit on noise alone.

Discriminator

A good discriminator should meet the following conditions when intended for color television and carrier telephone usage.

1. Frequency—amplitude characteristic must be linear over the entire i-f pass band.
2. Frequency—phase characteristic must be linear over the entire i-f pass band.
3. Conditions 1 and 2 must remain stable with respect to aging, temperature, power-supply variations, and mechanical shock and vibration.
4. Video output must be sufficient so that excessive video gain is not required.

The first condition specified can be met by making the bandwidth of the discriminator wider than the band pass of the i-f amplifier and utilizing a circuit that is inherently linear. The phase characteristic can also be made linear over the i-f pass band by the same considerations. Condition 3 requires stable components. This precludes the use of vacuum tubes in the circuit associated with the discriminator. Condition 4 can be met by supplying the discriminator with a relatively high level of i-f voltage.

An investigation was made into the various discriminators now in common use. No discriminator of conventional type met our requirements so, by a process similar to that taken in designing the intermediate frequency amplifier gain stages, there was evolved a readily adjustable discriminator with an extremely high degree of linearity. J. H. Clark is covering the complete description of this discriminator in a patent application. This discriminator is essentially two reactive networks derived by the Fosters Reactance theorem.

The actual circuit used may be found in Fig. 4(A) and the reactance plot for each network is to be found in Fig. 4(B). The resistance $R1$ and $R2$ are made equal to the driving point impedance of each network. This bridge circuit consisting of $R1$, $R2$ and the two networks is a special case in which the driving point impedance remains essentially resistive and constant at all frequencies, therefore the load on the limiter is not varied with frequency. By properly adjusting the location of the poles and zeros, the rectified output voltage of the two networks in combination is linear with frequency within the limits of the i-f pass band. The rectifiers used in conjunction with the discriminator are matched germanium diodes that exhibit high rectification efficiency and are stable with respect to temperature. The network itself is constructed of glass-invar piston capacitors with a very low temperature coefficient and are mechanically stable. The inductors are wound on subminiature low-loss coil forms. The coils are space wound and the turns held in place with low-loss coil dope. The Q of the coils at the operating frequency exceeds 200. The load presented to the networks by the diode rectifiers may be compensated for by the adjustment of the values of $R1$ and $R2$ and thus maintain a balanced bridge circuit in spite of the load produced by the rectifiers. With this circuit it is possible to obtain a linearity better than ±1.0%. Since the band pass of the discriminator exceeds the band pass of the i-f by 50% it is not necessary to adjust the discriminator each time the i-f band pass is adjusted.

The bridge-type discriminator easily meets conditions 1 and 2. Tests under actual operating conditions and in the laboratory indicate that condition 3 is met adequately with respect to aging, temperature and mechanical shock and vibration. The effects of power supply variations cannot enter into the operation of this circuit. The video output of the discriminator is about 0.5 volts peak to peak for the full 20-mc i-f band-pass sweep. A cathode follower is connected to the discriminator output. It's function is an impedance transformer to drive a short 93-ohm cable between the i-f chassis and the video-amplifier chassis.

It is desirable to operate the microwave system with the local oscillator above the carrier frequency in the upper end of the band and with the local oscillator below the carrier in the lower end of the band. This mode of operation minimizes image interference problems. Transforming from one mode of operation to the other inverts the video signal. The crystals used in the discriminator are of plug-in type and can be reversed. Reversing the crystals compensates for reversing the mode of operation.

Automatic Frequency Control (AFC)

It has been common practice in microwave video transmission systems to lock the receiver AFC to the sync pulse from the video signal. This practice resulted in d-c restoration of the signal and aided in the correction of distortion in the low frequency response of the system. With the advent of color transmission, video signal pre-emphasis became more common When pre-emphasis is used, the low frequencies are attenuated with a consequent over-shoot or spike being produced at the sync pulse. Since the sync pulse no longer has a flat bottom for the AFC to operate upon, this type of AFC tends to create more trouble than it cures. Tests on several types of microwave systems in our Plant have indicated that the average type of AFC is superior to the sync pulse reference type when transmitting a pre-emphasized signal. For this reason an average-type frequency control is used. Optimum color operation often depends on precise adjustment of i-f operating frequency through adjustment of the AFC, therefore it should be easily adjustable.

It is possible to use the video discriminator network as the AFC discriminator network. For this purpose an additional set of diodes is connected to the video discriminator. The output of these diodes is connected to a potentiometer as indicated in Fig. 4. The output voltage of the potentiometer is connected to a bridge circuit that controls the negative voltage applied to the repeller of the klystron local oscillator. This system of feed-back local oscillator control operates on the klystron to bring the voltage output of the potentiometer to zero. Since the potentiometer can vary the sample being taken from each network, the fre-

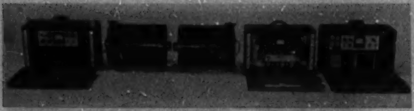


Fig. 6. Front view of transmitter control, transmitter receiver, receiver control, and receiver power supply. Covers off in operational condition

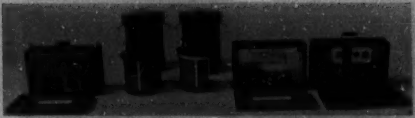


Fig. 8. Same equipment as Fig. 6. Rear covers removed for minor maintenance testing. Safety shield in place

Fig. 7. Same as Fig. 6. Rear view

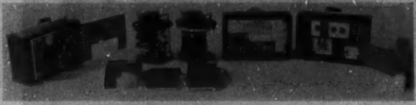


Fig. 9. Same view as Fig. 6. Rear covers and safety shield removed for major maintenance testing

quency of the carrier in the i-f amplifier can be varied by adjusting the potentiometer.

This method of controlling AFC has a high order or resetability and is very easily adjusted. The gain in the AFC feed-back loop determines the degree of AFC stability achievable. In this circuit, a carrier frequency drift of 10 mc can be held to within 500 kc at the i-f amplifier channel.

Video Amplifier

A video amplifier must be capable of raising the voltage level of the output of the discriminator to at least 1.5 volts peak to peak into 75 ohms without introducing differential phase or gain distortion. Means must be included to set the gain of the amplifier so the peak to peak output voltage can be adjusted. The usual low-impedance output circuits were investigated and found lacking in one or more of the requirements. Basically, any amplifier device where the output voltage depends on large percentage of variations in the plate current of the vacuum tube is subject to large values of differential phase and gain distortion. The differential gain is due to the difference in transconductance of the vacuum tube at various instantaneous plate currents. The differential phase is caused by the difference in the resistance-capacitance time constant introduced by the output capacity and the varying plate resistance of the tube. A circuit was tried that has been previously used mainly for its ability to have both a fast rise and a fast fall time. This is a combination cathode and anode follower herein referred to as a cathanode follower. The cathanode follower as shown in Fig. 5 is characterized by very small changes in plate current causing relatively large changes in output voltage. This is because the tubes are driven out of phase and as the plate resistance of one tube goes up the plate resistance of the other tube goes down resulting in an almost constant plate current.

The output impedance of the video amplifier is made very low (10 ohms) by application of a high percentage of feedback around the amplifier. A series resistance is connected between the output of the amplifier and the coaxial output connector to match the characteristic impedance of the coaxial cable. The effect of nonlinear output impedance in the video amplifier is minimized because the greater part of the output impedance of the amplifier is a resistor. With this design it is possible to obtain multiple outputs by providing additional series resistors connected to the other output connectors. The only effect that the additional outputs have on each of the other outputs is that the maximum permissible voltage drive is reduced. As long as the output voltage of the amplifier is less than the maximum permissible, the connection of additional lines does not influence the output voltage on the other lines. On this system, two outputs and one high quality monitor point are provided. The high quality monitor point has the same voltage output into high impedance that the other outputs have into 75 ohms. The output impedance of the monitor point is 75 ohms, therefore the length of the cable connecting it to a high-impedance oscilloscope will not affect the presentation until the length of the cable is such that the transmission characteristic of the cable itself becomes a significant factor.

Three stages of voltage amplification precede the output stage. The first stage is not within the feedback loop and has the function of equalizing the loss in the gain frequency characteristic introduced by the discriminator and coupling coaxial cable. This equalization is not adjustable. The second and third stages are within the feedback loop. Small changes in the gain frequency characteristic that may be necessary can be made up by adjustment of the feedback capacitor.

With the lumped equalization removed, this video amplifier can be made flat from 10 cycles to 10 mc with the frequency response continuing to 14 mc before dropping below −6 db. The video amplifier is

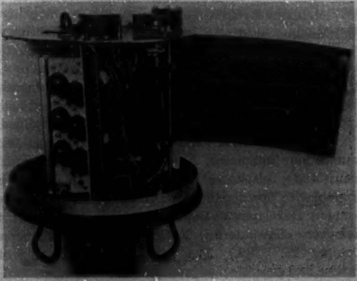


Fig. 10. Receiver unit top view. Safety shield removed to illustrate component placement. The elements to the right of center bulkhead present contact hazards. No contact hazards exist on left of center bulkhead where all system adjustment points are located

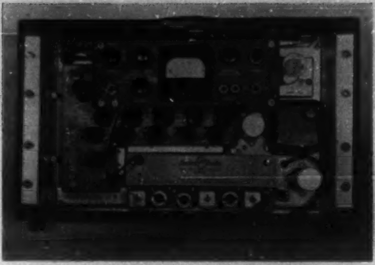


Fig. 11. Receiver control unit. Front view

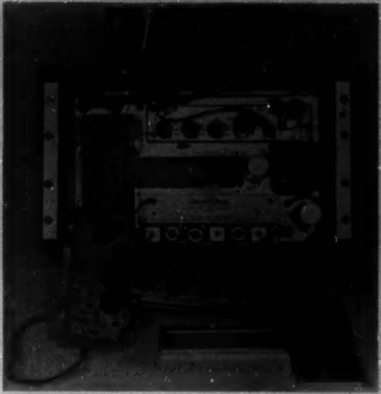


Fig. 12. Receiver control unit with I-f amplifier removed from main chassis. A power extension cord permits operation of I-f amplifier in this position

capable of 6 volts peak-to-peak output into a single 75-ohm line and 3 volts peak to peak onto 2 lines at a differential phase and gain of less than 1/4 of a degree and 1/10 of a db. This gives a 3 to 1 margin for feeding 2 lines at one volt which is the normal operating procedure. Sixty-cycle voltages that are introduced into the amplifier due to stray common ground paths can be cancelled out by means of a 60-cycle balance control which injects a correctly phased 60-cycle voltage directly into the cathode circuit of the cathanode follower. By this means the 60-cycle output of the receiver can be reduced to 1.5 millivolts rms.

System Operational Characteristics

By operational characteristics is meant the manipulative techniques required to install, align, and operate the overall relay system. The first and guiding principle observed was of course the safety of personnel. This was considered from two aspects. The first was, given an operable system, how it could be made safe for all personnel who were required to operate it. The second aspect was, given the requirement to do alignment or repair work, how the system could be made safe for people to do this work even though the normal operator protection had been removed.

When faced with such an important and complex requirement it is helpful to use the results of recognized authorities when searching for basic principles to apply. ("Protection of Personnel from Voltages Higher Than 150 Volts in Communication Equipment, Testing Equipment and Telephone Power Plants." Unpublished Specification for Equipment Systems Development Department, Bell Telephone Laboratories, X64587 dated October 26, 1949.) The most clear, concise, and general set of principles found were these: insulate, isolate, guard, and warn.

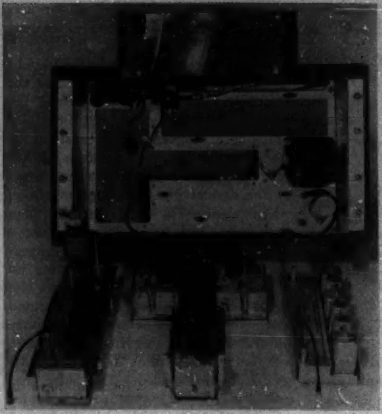
In all cases where an accidental contact with high potential seemed possible the point in question was insulated. High-potential leads were routed through normally inaccessible locations. Safety-guard shields which could be removed and replaced simply were devised for all points where hazardous potentials were accessible. Adequate warning signs and schematic signs were applied at all points when it was felt the accident possibilities could be reduced. Figs. 6–10 illustrate the mechanical arrangements evolved to implement the personnel safety objectives for the system. It should be noted that the safety shielding is readily mounted or demounted by maintenance personnel and it is felt that this feature will insure that shielding will be used by maintenance personnel when it is feasible to do so during the course of maintenance operations.

By being careful in the design of all of the active elements of the system it was possible to minimize system equalization procedures. Equalization for a monochrome television signal consists of correctly aligning the system according to a simplified maintenance and operation instruction. Equalization for a NTSC color signal requirement demands a little more care in shaping the system amplifier response curves. An exceptance requirement for the system is that the simultaneous requirements of flat band-pass to 8 mc, differential phase below 3 degrees and differential gain below 0.5 db can be met in this manner. Once a system is optimized by these procedures, pre- and de-emphases may be applied to the video portion of the circuit further to improve the system differential phase and differential gain performance.

An attempt was made to simplify the system control arrangement. Metering and operational controls were conveniently arranged on the front panel of the receiver control unit and minor field maintenance adjustment points located in a readily accessible positions. Considering that a continuing training responsibility falls on supervisory personnel responsible for the operation of this equipment a comprehensive operating instruction book was prepared. It was the objective that any person with a minimum of knowledge of microwave radio relay equipment could successfully operate the system with no instruction other than reading this operators instruction pamphlet. It is felt that the operation instructions materially increase the use possibilities of the system in that any licensed person can, unaided, establish a satisfactory circuit if he but follow the printed directions.

System Maintenance Characteristics

When meeting a continuing demand for quality broad-band transmission facilities on a routine basis the facet of the operation concerned with system maintenance assumes significant importance. Considering that the various units are portable and that any one of 40 or more



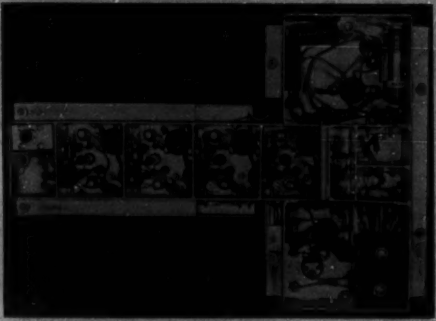


Fig. 13 (left). Receiver control unit. Principal subcomponent chassis removed. Subcomponent chassis are video amplifier, i-f amplifier, AFC circuit

Fig. 14 (above). Main i-f amplifier, limited, and discriminator with rear cover removed, showing parts layout, interstage shielding, and measures taken to prevent entry of interfering signals

possible operators may be assigned to transport, install, operate and clear trouble on any one of the 30-odd units in plant the adoption of certain standard maintenance objectives is of paramount importance. The standards were set forth in a comprehensive manual written to be readily understood by maintenance personnel. The utility of this manual was checked by a field trial and the recommendations of maintenance personnel incorporated in its final version.

The complete microwave system consists of five principal units as shown in Fig. 6. These units are designated: transmitter control unit, transmitter unit, receiver unit, receiver control unit, and receiver power supply unit.

The circuitry of the transmitter, transmitter control and receiver power supply units was considered adequate. It was was necessary however completely to redesign the receiver and receiver control units. The subunits of these principal units were designed on demountable subchassis assemblies. With subchassis construction it is possible to demount all principal electronic subassemblies for maintenance purposes. This feature materially reduces trouble clearing time on the relay units in that spare subcomponents chassis may be aligned and held for installation in systems which demonstrate trouble in the respective components. The possibility of separate alignment of subcomponents materially reduces system maintenance outage time. Figs. 11–13 illustrate the utility of plug in subcomponent construction. Fig. 14 shows the main i-f amplifier, limiter, and discriminator.

Experience in the maintenance of systems designed to operate at video and intermediate frequencies indicated that all such units should be aligned with all covers in place in service condition. Consequently, all adjustments on the various subassemblies were located with components positioned for tuning with the subunit installed for operation. It is not necessary to remove subcomponent covers to perform routine maintenance tests or adjustments. This feature enables the maintenance man to optimize a system in service condition and effectively eliminates the heretofore necessary step of rechecking after maintenance adjustment have been made and covers replaced on the various active elements.

Each stage must be individually aligned so one stage does not compensate for nonlinearity in a preceding stage. To this end, each i-f amplifier stage is equipped with a test point. Each test point contains its own detector circuit, so it is only necessary to connect an oscilloscope to the test point when making gain and band pass tests; see Fig. 1. The test point itself is a Microdot subminiature coaxial connector. A capacitor is mounted at the connector to prevent entry of interfering signals into an otherwise well-shielded i-f enclosure. Tube failures can be quickly located in the field by measuring the signal voltage at each stage using an ordinary multimeter. In addition to band-pass and trouble-location tests, these test points may also be used in the measurement of receiver noise figure, conversion loss, and the recording of signal levels for engineering studies of microwave paths.

Conclusions

It is possible to extend economically the useful life of certain obsolescent microwave radio relay equipment by the redesign of critical circuitry. NTSC color television signals may be transmitted on portable microwave radio relay units in a routine manner if an adequate system is available and standardized system operational and maintenance instructions are provided.

The iteration of identical i-f amplifier coupling networks tends to reduce the variation of system differential phase and differential gain with tube changes and path changes. The application of plug-in subcomponent chassis construction techniques materially reduce system maintenance expense.

Field tests of the completed system indicate video base band equalization is less time consuming and technically superior to i-f amplifier equalization when correcting differential phase and differential gain difficulties.

References

1. Differential Phase and Gain Measurements in Color Television Systems, H. P. Kelly. *AIEE Transactions (Electrical Engineering)*, vol. 73, pt. I, Nov. 1954, pp. 565–69.

2. One Approach to a Video SHF Relay System, R. H. Coe, F. F. McClatchie. *Ibid.*, vol. 73, pt. I, Sept. 1954, pp. 323–29.

The Response of Resistance Voltage Dividers to Steep-Front Impulse Waves

A. F. ROHLFS
MEMBER AIEE

J. S. KRESGE
ASSOCIATE MEMBER AIEE

F. A. FISHER
ASSOCIATE MEMBER AIEE

Synopsis: At present there is no generally accepted technique for determining the response of high-voltage dividers to fractional microsecond impulse waves. This paper describes a method which uses a low-voltage cable-type pulse generator and a high-speed oscilloscope as the basic tools. A technique for calibrating the oscilloscope amplifier and the associated input attenuators is described. The effects of such parameters as total divider resistance, divider length, and connecting lead lengths have been investigated. The accuracy and limitations of vertical resistance dividers for measuring steep-front waves are evaluated.

HIGH-IMPULSE voltages are generally measured by means of a cathode-ray oscillograph and a voltage divider, with resistance, capacitance, or mixed resistance-capacitance (*R-C*) dividers all being employed. The accuracy of such a system is determined primarily by the response characteristics of the divider since the oscillograph introduces negligible wave-shape distortion and can be easily calibrated.

The response of high-voltage dividers has received attention by several investigators, but much of the published work is based upon theoretical calculations with little experimental verification. In addition, the response to steep-front waves has to a large extent been neglected, although the amplitude errors for such waves are generally greater than for full waves or waves chopped on the tail. For the purposes of this paper a steep-front wave is defined as a linearly rising wave which chops abruptly to zero at crest time. In some of the literature such a wave is referred to as a front-chopped wave.

The present investigation has a two-fold objective:

1. To develop an experimental technique for checking the performance of a divider system as a whole. The divider system is taken to include the leads from the divider to the test object and the deflection cable from the divider to the oscillograph deflection plates.
2. To investigate the limitations of presently used resistance dividers.

Experimental Technique

Other investigators have used one or more of the following techniques for evaluating divider systems.

1. Theoretical calculations based upon known or assumed parameters, generally with a few simplifying assumptions about the effects of some parameters.[1–4]
2. Sustained frequency response tests usually extending well into the megacycle frequency region.[4–6]
3. Comparison against a standard device such as a sphere gap or a standard divider.[7]
4. Tests with low-voltage impulse waves using a high quality amplifier to amplify the divider output wave for direct comparison with the input wave. Such tests may involve a comparison of input and output wave forms, amplitudes, or both.

Theoretical calculations are generally necessary to gain a thorough understanding of the principles involved, but reliance on calculations alone without experimental verification can be misleading. This is particularly true when, in problems such as this one, certain assumptions must be made both as to the magnitude of some of the parameters and their effect on the over-all system.

Sustained frequency-response tests are a valuable aid for making qualitative comparisons between several dividers. However, to determine the impulse response of a given divider from the frequency response a calculation involving a Fourier integral transformation is required.

Comparison against a standard device has the obvious disadvantage that one does not know the quality of the standard without testing it. If the standard could be tested then the same test would suffice to check the divider in question. Tests with low-voltage impulse waves and an amplifier are feasible provided the amplifier response is much faster than the system being tested. Amplifiers with rise times as fast as 0.01 μsec (microsecond) are now commercially available and it can be demonstrated that they are suitable for testing dividers with waves cresting at 0.05 μsec or longer.

With due consideration to the various possible methods it was decided that the impulse-amplifier technique offered the most direct method of experimentally checking a complete divider system. A testing technique yielding data on the relative amplitude as well as wave shapes of the divider input and output waves was devised. The basic tools for these tests were a high-speed oscilloscope and a square-wave pulse generator. The oscilloscope employs an amplifier having an over-all rise time of 0.012 μsec. The input circuit consists of compensated attenuators giving a deflection sensitivity on the screen ranging from 20 to 0.05 volt/cm. Sweep rates as fast as 0.02 μsec/cm. are provided. The pulse generator was a cable-type generator employing 75-ohm coaxial cable discharged through a mercury wetted-contact relay. Two output circuits are provided, each with a 75-ohm ladder-type attenuator which can introduce up to 84 db (decibels) attenuation.

Calibration of Measuring Circuits

Since comparisons of the amplitude of the input and output voltages of the divider were to be made it was necessary to determine the relative sensitivity of the various input attenuator positions of the oscilloscope. Basically this was done by applying two voltages differing by a known ratio and observing the deflection caused by each of these voltages. Deferring for the moment the problem of determining the true voltage ratio, the procedure is this:

1. Apply a high voltage V with the input attenuator set on a low sensitivity position S and observe the oscilloscope deflection Y.
2. Apply a low voltage V' with the input attenuator set on a high sensitivity position S' and observe the oscilloscope deflection Y'.

The sensitivity ratio for the two attenuator positions is then as given by the following equation

$$\frac{S}{S'}=\frac{V}{V'}\frac{Y'}{Y} \tag{1}$$

The two voltages, V and V', were obtained by introducing attenuation with the output attenuator of the pulse generator and were measured by means of a d-c bias voltage using the circuit shown in Fig. 1. The operation of the circuit is as follows:

1. With no pulse applied to the system adjust potentiometer P_2 until V_2 equals

---

Paper 57-353, recommended by the AIEE Special Instruments and Auxiliary Apparatus Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 20, 1957; made available for printing May 17, 1957.

A. F. ROHLFS, J. S. KRESGE, and F. A. FISHER are with the General Electric Company, Pittsfield, Mass.

The authors wish to acknowledge the guidance and valuable suggestions of J. H. Hagenguth.

Vol.
Pt. 1
195

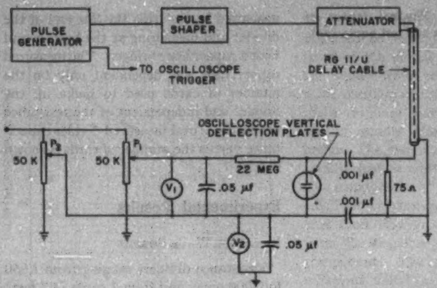

PULSE GENERATOR
PULSE SHAPER
ATTENUATOR
TO OSCILLOSCOPE TRIGGER
RG 11/U DELAY CABLE
OSCILLOSCOPE VERTICAL DEFLECTION PLATES
50 K
P_2
50 K
P_1
V_1
.05 μf
22 MEG
.001 μf
.001 μf
75 Ω
V_2
.05 μf

Fig. 1. Circuit for calibration of pulse generator attenuator

the average voltage on the sweep plates in order to avoid defocusing of the electron beam. Then set potentiometer P_1 so that V_1 equals zero and mark the reference position of the beam.

2. With pulse voltage V applied adjust potentiometer P_1 until the crest of the wave is at the reference position marker. The d-c voltage V_1 is then equal to the crest amplitude of V.

3. With a different pulse voltage V' applied readjust P_1 to bring the crest of the wave to the reference marker. V_1' is then equal to the crest amplitude of V'.

The ratio of the two voltages is then

$$\frac{V}{V'} = \frac{V_1}{V_1'} \qquad (2)$$

Having established the ratio of the voltages, equation 1 can be solved for the relative sensitivity of two attenuator positions. The output of the attenuators was also examined for wave-shape distortion and for the waves used in this investigation the distortion was negligible.

Tests on Dividers

The circuit used for checking high-voltage dividers is shown in Fig. 2. When the charged generator cable is discharged into the matching resistor network a square pulse of length equal to twice the travel time of the generator cable will be developed at point A; see Fig. 3(A). This pulse is then modified as desired with the pulse-shaping network and inverter cable shown. If the time constant of the pulse-shaping network (charging through 75 ohms) is long compared to the length of the voltage pulse at point A the voltage developed at point B will be nearly a symmetrical triangular wave as shown in Fig. 3(B). A pulse approximating the ideal steep-front wave will be developed if at time t_1 a square wave of proper amplitude, but of opposite polarity to the wave at A is injected at point B. The inverted wave is obtained by using an inverter cable as shown in Fig. 2. The resulting wave at point B will then have a nearly linear rate of rise up to time t_1 and a rapid fall to zero at t_1 see Fig. 3(C). The wave shape between times t_1 and t_2 can be kept near zero by the proper choice of pulse shaping network parameters and generator cable lengths. The time at which the inverted pulse is applied, that is the chop time, depends on the length of the inverter cable while the amplitude of the inverted pulse depends on the amount of attenuation introduced by attenuator B.

Pulses so generated were applied to the top of the test divider through a sloping lead. Since the line lead should be considered as part of the voltage divider the input voltage was measured at the input end of the lead. The divider input and output voltage waves were passed through equal lengths of 75-ohm coaxial cable to the oscilloscope to eliminate the effects of all measurement time delays except those inherent in the divider circuit.

The divider input and output waves were both applied to the input of the oscilloscope and the calibrated attenuator was used to obtain the proper deflection on the screen. The input attenuators were used to keep the voltage applied to the amplifier itself nearly the same, whether measuring the divider input or the response wave. In this manner the gain of the amplifier was held constant so that any slight amplifier distortion would affect both the input and response waves and so tend to be nullified. When measuring the divider output wave the divider input measuring cable was left connected, but was terminated at the oscilloscope end with the same values of resistance and

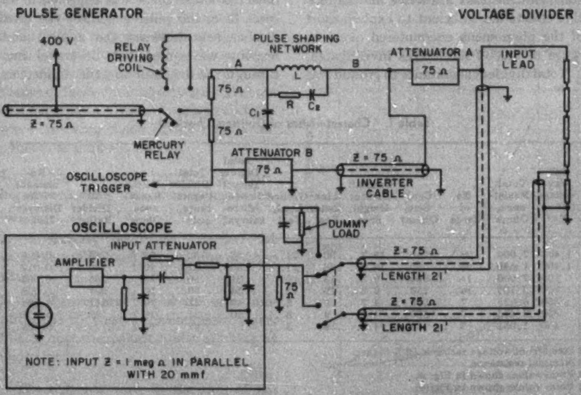
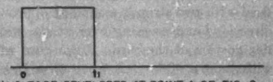
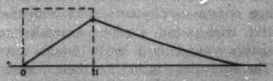

PULSE GENERATOR
VOLTAGE DIVIDER
400 V
RELAY DRIVING COIL
PULSE SHAPING NETWORK
A
B
L
ATTENUATOR A
75 Ω
INPUT LEAD
75 Ω
75 Ω
C_1
R
C_2
Z = 75 Ω
MERCURY RELAY
ATTENUATOR B
75 Ω
Z = 75 Ω
INVERTER CABLE
OSCILLOSCOPE TRIGGER
OSCILLOSCOPE
INPUT ATTENUATOR
AMPLIFIER
75 Ω
DUMMY LOAD
Z = 75 Ω
LENGTH 21'
Z = 75 Ω
LENGTH 21'
NOTE: INPUT Z = 1 meg Ω IN PARALLEL WITH 20 mmf

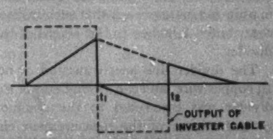
Fig. 2. Circuit for calibration of voltage dividers



(A) VOLTAGE DEVELOPED AT POINT A OF FIG. 2



(B) VOLTAGE DEVELOPED AT POINT B IF INVERTER CABLE IS NOT CONNECTED


t_1
t_2
OUTPUT OF INVERTER CABLE

(C) VOLTAGE DEVELOPED AT POINT B WHEN INVERTER CABLE IS CONNECTED

Fig. 3. Voltages at points A and B of Fig. 2

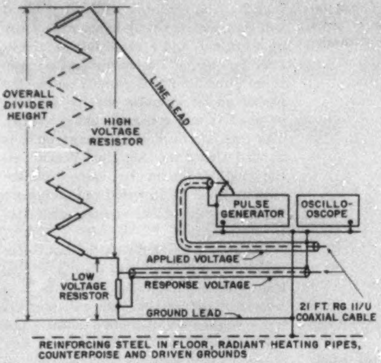


Fig. 4. Physical layout of dividers and test equipment

capacitance presented by the oscilloscope. This was done to insure that the voltage applied to the divider was not changed when the divider input cable was disconnected from the oscilloscope.

Types of Dividers Tested

The dividers tested were vertical resistance dividers of the same construction as in presently used by the authors' company for both laboratory and commercial impulse measurements. A tabulation of the dividers tested, giving the pertinent physical and electrical characteristics may be found in Table I. The high-voltage resistance elements were noninductive wire-wound resistance cards with the winding occupying approximately 8 inches of a 10- by $2^1/_2$- by 3/32-inch winding form. A double-winding technique was used with two strands wound in opposite directions and crossing over at the wide flat portion of the card. Each card will withstand a 50-kv full wave or a 75-kv steep-front wave without experiencing excessive temperature rise. The low-voltage resistance elements were noninductive wire-wound vitreous enameled resistors wound on 1 by 9/16-inch cores. The equivalent rating for these resistors would be $12^1/_2$ watts. The high-voltage resistors were mounted in a vertical rack in a zig-zag manner such that each resistor contributed 8 inches to the height of the divider.

The divider line lead consisted of no. 16 copper wire, and the ground end of the divider was connected to the common ground system of the measurement circuit with 2-inch-wide aluminum foil. The effective ground plane for the dividers consisted of the bonded reinforcing steel and radiant heating system in the building floor. All vertical ground planes, such as building walls, were at least two divider heights removed from the divider. A sketch of the physical layout of the test circuit is shown in Fig. 4.

Divider Parameters

The residual inductance, and for some cases the effective resistance of the cards and the low-voltage resistors were measured with a radio-frequency bridge. The L/R (inductance/resistance) ratio of all high-voltage cards used is shown on Fig. 5. The inductance of each divider tested was obtained from this figure. The effective resistance as a function of frequency for several cards is shown on Fig. 6. While the resistance increases with increasing frequency, the experimental results indicate that inductive and capacitive effects are sufficient to explain most of the phenomena encountered over the range of steep-front waves investigated.

Total divider capacitance to ground was measured by isolating the line end of the divider and measuring at the bottom end with a capacitance bridge. The measured capacitance was dependent only on the number of cards used to make up the divider and independent of the resistance per card. Total measured divider capacitance versus the number of cards is shown in Fig. 7.

Experimental Results

Response Wave Shapes

Resistance dividers ranging from 1,050 to 7,000 ohms and from 7 cards (4.7 feet) to 28 cards (18.7 feet) were tested. Three types of waves were used: a square wave with a front of 0.002 μsec and a duration of 1.5 μsec, steep-front waves having a linearly rising front and a sharp cutoff, and triangular waves having a linearly rising front and linearly decaying tail. The front time of the square wave was measured by direct insertion to the oscilloscope plate system without using the amplifier. Oscillograms of the various applied waves and the corresponding divider response waves are shown in Figs. 8 through 11. These oscillograms are grouped to show the effect of divider length, divider resistance and wave shape. Fig. 8 shows the response of 7,000-ohm dividers having different lengths, Figs. 9 and 10 show the response of dividers of a fixed length but different resistances, while Fig. 11 shows the effect of wave shape.

The wave applied to the divider line lead and the response wave measured at the oscilloscope end of the deflection cable were recorded on the same oscillogram. Since the oscilloscope was triggered in all cases from the pulse generator the time displacement between the applied and response waves represents the travel time through the line lead and the divider, as

Table I. Characteristics of Dividers Checked

Full-Wave Voltage Rating, Kv*	Total Resistance, Ohms	No. of Cards	Card Size, Ohms†	Over-all Divider Height, Feet	Line Lead, Feet	Ground Lead, Feet	Total Divider Inductance, Microhenrys‡	Total Divider Capacitance, μμf§	Low Voltage Resistance, Ohms¶	Divider Ratio**	Resistance Wire Diameter, Inches
1,400	7,000	28	250	18.7	30	21	128	82	76.7	185.5	0.007
1,400	1,680	28	60	18.7	30	21	80	82	10.1	190	0.012
700	7,000	14	500	9.3	15	12	81	50	76.7	185.5	0.005
700	2,105	14	150	9.3	15	12	54	50	10.1	237	0.008
350	6,825	7	1,000	4.7	9	8	40	31	76.7	181	0.004
350	3,495	7	500	4.7	9	8	40	31	76.7	93.8	0.005
350	1,053	7	150	4.7	9	8	27	31	10.1	119.6	0.008

* Steep-front voltage rating is 50% higher.

† Nominal resistance.

‡ From values shown in Fig. 4.

§ From values shown in Fig. 6.

¶ Noninductively wound on a 1- by 9/16-inch core.

** $\frac{R_1}{R_e}+1$ where R_e = low-voltage resistor in parallel with 75 ohms.

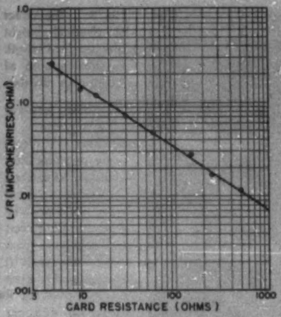


Fig. 5. L/R ratios of resistance cards

well as the distortion caused by the circuit parameters of the divider. Since the oscilloscope sensitivity ratio could not be changed by the same factor as the divider ratio, the response waves shown on the oscillogram are not directly comparable. Before comparing amplitudes the response wave height should be multiplied by the correction factor n which appears on each oscillogram where

$$n=\frac{\text{divider resistance ratio}}{\text{oscilloscope sensitivity ratio}}$$

WAVE COMPONENTS

It is evident that the divider distorts the applied wave shape to some degree, but in most cases the response wave crests at or only slightly later than the crest of the applied wave. In some cases where the response wave does not reach its maximum amplitude until later there is at least a dip in the response wave at the crest time. A particularly vivid example is seen in oscillogram 558, Fig. 8. The shapes of these various response waves are best explained by visualizing a steep-front wave as being made up of three components as shown in Fig. 12.

1. A linearly rising portion *1A* starting at zero time and continuing indefinitely.

2. A linearly descending portion *2A* with the same slope, but starting at chop time.

3. A square wave *3A* of the same amplitude as the crest of the composite wave, but of opposite polarity, also starting at chop time.

The response wave is the sum of the responses to these three components as shown at the bottom of Fig. 12. The response to the two linear portions was obtained by extrapolating the first portions of the actual response wave while the square-wave response was obtained from oscillogram 563. The sum of the three response waves is a wave with a dip or saddle after the applied wave crests. The response wave does not drop rapidly just at crest time since the negative components of the response wave are small for a certain period of time after the true crest and so subtract only slightly from the response to the linearly rising component. The saddle and other oscillations noted on oscillograms 558 and others are due to the oscillating response to the square-wave component. This oscillating square-wave response is a manifestation of the inductance and capacitance of the divider system.

EFFECT OF LINE LEAD

Fig. 8 shows the response of a 28-card 7,000-ohm divider with and without a 30-foot line lead. The response with no line lead was measured by bending the divider and bringing the line end back down to the pulse generator. This changed the capacitance characteristics of the divider slightly but the results at least indicate the approximate effects of the line lead.

It will be noted that with the line lead the response wave crests approximately 0.03 μsec after the applied wave, which is the time required for the wave to travel through the 30-foot line lead. Without the line lead, the first crest of the response wave occurs at the same time as the applied wave crest. This time delay introduced by the line lead does not in itself distort the response wave since all portions of the wave are similarly delayed. However, an oscillation is set up between the inductance of the lead and the effective line end capacitance of the divider. The period of this oscillation for the 28-card divider is 0.15 μsec. Assuming the line lead to be 15 microhenries (0.5 microhenry per foot) the effective line-end capacitance of the divider is 38 μμf (micromicrofarad) compared to the total divider capacitance to ground of 82 μμf. The effect of the line-lead oscillation on the magnitude of the response wave will be discussed later.

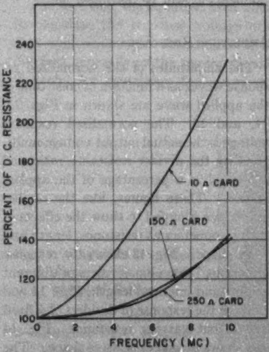


Fig. 6. Effective resistance of resistance cards

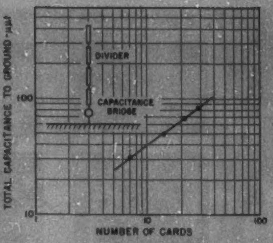


Fig. 7. Capacitance to ground of dividers

INTERNAL OSCILLATIONS

Oscillations in the response wave may also be the result of the divider oscillating within itself. Fig. 10 shows that as the resistance of the 7-card divider is decreased, the oscillations of the square-wave response increase. Since the line-lead inductance and divider line-end capacitance remained the same, the increased magnitude of the oscillations must be due either to oscillations within the divider or to decreased damping of the line-end oscillation by the divider resistance. The latter does not appear to be likely since the divider resistance acts in parallel with the line lead inductance and line-end capacitance for damping. Thus when the divider resistance was decreased, the oscillations would have decreased.

An equation for the response of a divider having resistance and inductance in series and capacitance to ground is given by J. H. Hagenguth[3] for a square applied wave. From it the critical damping resistance for the fundamental oscillation is

$$R=2\pi\sqrt{\frac{L}{C}} \qquad (3)$$

and the period of oscillation is

$$T=2\sqrt{LC} \qquad (4)$$

Calculated periods, critical damping

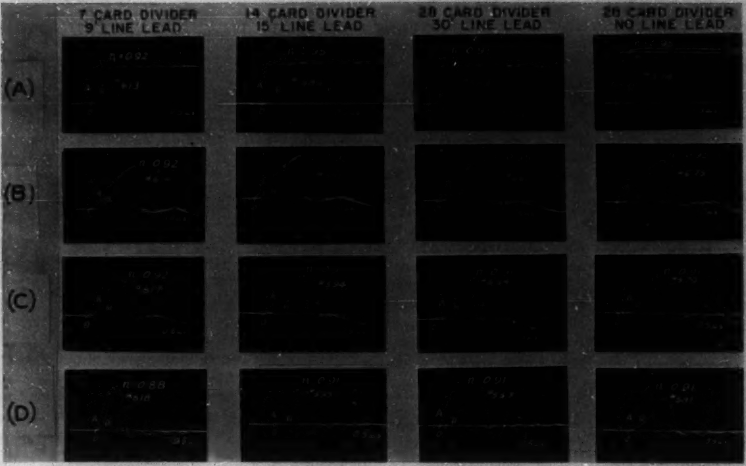


Fig. 8. The effect of physical height and line-lead length on the response of 7,000-ohm dividers

A—Square - wave response
B—Steep - front wave response 0.51 microsecond to crest
C—Steep - front wave response 0.19 microsecond to crest
D—Steep - front wave response 0.10 microsecond to crest
Note: Multiply R by n to compare amplitudes in Figs. 8–10

resistances, and the measured periods are as given in the table following:

	7-Card Divider, 7,000 Ohms	7-Card Divider, 1,050 Ohms
Calculated period, μsec	0.08	0.06
Calculated damping resistance, ohms	7900	5800
Measured period, μsec	0.05	0.07

These calculations indicate that the 7,000-ohm divider should be essentially damped, yet the square-wave response is oscillatory. Since the calculated period of oscillation does not agree with the measured value, 0.08 versus 0.05 μsec, the oscillation of the 7,000-ohm divider is probably due to the line-lead inductance. Using 4.5 microhenries for the line lead inductance and 14 $\mu\mu$f for the line-end capacitance of the divider the period of oscillation is calculated to be 0.05 μsec which is the same as the measured value again indicating that the oscillation is due to the line lead. The capacitance value of 14 $\mu\mu$f was obtained by multiplying the total capacitance (31 $\mu\mu$f) by the ratio of line end capacitance to total capacitance calculated for the 28-card divider

$38/82 = 0.46$

The calculations also indicate that the 1,050-ohm divider is underdamped and that the resulting period of oscillation should be 0.06 μsec. Since the measured period is 0.07 μsec, a reasonably close check, this oscillation is believed to be mainly internal. Part of the observed oscillation is of course due to the line lead, but this should have the same period as, and should be more nearly damped than, the oscillation noted on the 7,000-ohm divider.

It is evident that oscillations in the response wave may be due to either the inductance of the line lead oscillating with the line-end capacitance of the divider or oscillations built up within the divider itself. The former may be controlled by keeping the line lead short; the latter by maintaining sufficient resistance in the divider to prevent oscillation. However, this is in opposition to the requirement of using low resistance to obtain a low R-C constant necessary for short time measurements as will be discussed later.

AMPLITUDE RESPONSE

The amplitudes of the normalized response waves as a function of crest time of the applied wave are shown in Figs. 13, 14, and 15. The normalized response voltage is the actual output voltage multiplied by the divider resistance ratio and expressed as a percentage of the applied voltage. These figures, like the oscillograms, are grouped to show the effects of such parameters as resistance, length and wave shape. Fig. 13 shows the response of various 7,000-ohm dividers of different heights and line-lead length, Figs. 14 and 15 show the response of dividers of a fixed length, but varying resistance. Fig. 14 also shows the effect of wave shape. The measured response voltages are shown as points connected with straight lines. For response waves with a saddle and a second crest, as in oscillogram 558, the response voltage for each of the crests is shown. The points representing the voltage at the true crest time are connected by dashed lines; the points representing the crest with the largest amplitude are connected by solid lines.

Griscom, Lloyd, and Hileman[4] have derived an expression for the response to a linearly rising voltage for a pure resistance divider having only stray capacitance to ground. As shown in the appendix, this expression may be simplified to the following equation which gives the error in the response wave at crest time (t) of an applied steep-front wave.

$$\text{Per cent error} = \frac{RC}{6t}\,100 \qquad (5)$$

when $RC \leq 2t$ and where

R = total divider resistance
C = total divider capacitance to ground

The calculated response voltage using the foregoing equation is shown for each divider as a smooth solid curve. In almost all cases the measured error of the response voltage is less than that predicted by equation 5. There are several reasons for this:

1. The applied waves used during the investigation round over slightly near crest instead of continuing to rise linearly. Such wave shapes will allow the response voltage to rise nearer to the true value than would a linearly rising wave which is sharply cut off.
2. Oscillations resulting from the induct-

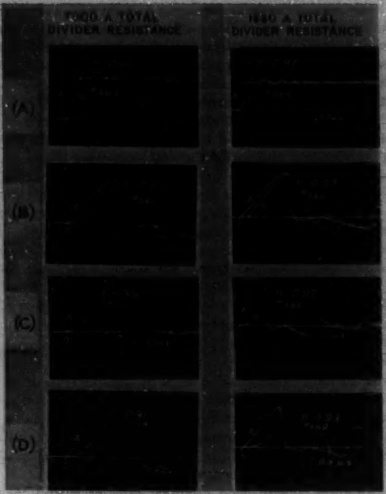


Fig. 9. The effect of over-all resistance on the response of 28-card dividers

A, B—Same as in Fig. 8
C—Steep-front wave response 0.26 microsecond to crest
D—Steep-front wave response 0.19 microsecond to crest

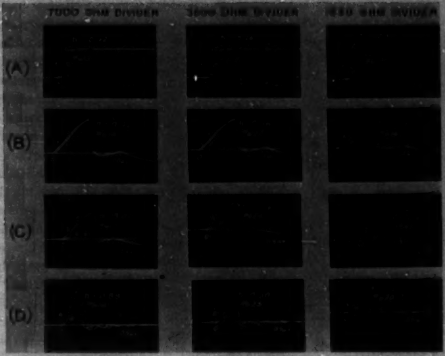


Fig. 10. The effect of over-all resistance on the response of a 7-card divider with a 9-inch line lead

A, B, C, D—Same as in Fig. 8

ance of the line lead and in the divider itself may increase or decrease the response voltage depending upon the period of oscillation and the time at which the applied wave is chopped. It is evident, for example, that the oscillations due to the line lead of the 28-card divider increased the response voltage above that obtained without a line lead. See Fig. 13.

3. The characteristics of the low-voltage resistor across which the response voltage is measured can have a pronounced effect on the response. Ideally it should have the same *RLC* characteristics per ohm as the main portion of the divider. If the inductance of the low-voltage resistor is too large, the voltage will be greater than otherwise or if the capacitance is too large the voltage will be low. Time did not permit investigation of these effects beyond establishing that the resistors used were not excessively inductive or capacitive.

Effect of Wave Shape

It should be emphasized that these results apply to steep-front waves with a rapid chop off. A divider will reproduce a symmetrical triangular wave more accurately than a steep-front wave. This is because the triangular wave does not have a negative square wave component such as shown in Fig. 3(C). The absence of this component allows the response wave to continue to rise after the applied wave has crested and to reach more nearly the correct magnitude. The improved accuracy is clearly indicated in the oscillograms of Fig. 11 and the normalized response curves of Fig. 14.

Accuracy of Vertical Resistance Dividers

The fact that the equation used above to calculate the response voltage error does approximate the actual error of the dividers, generally on the pessimistic side, is useful for examining the range over which vertical resistance dividers may be used.

The relationship between capacitance and divider length is shown in Fig. 7 and may be approximated as 3 μμf per resistance card for the longer dividers. Since one card will withstand at least 75 kv for short duration waves the capacitance per kilovolt for a divider being used at its maximum voltage is 0.04 μμf. Thus the capacitance of a divider of minimum length for the voltage being measured is

$$C=0.04V_{kv}(10^{-12})\text{ farads} \tag{6}$$

Substituting this factor in equation 5 gives

$$\text{Per cent error}=0.67R\frac{V_{kv}}{t}10^{-12} \tag{7}$$

This may be written in a more general form

$$\text{Per cent error}=\frac{cR}{6}\frac{V_{kv}}{t}100 \tag{8}$$

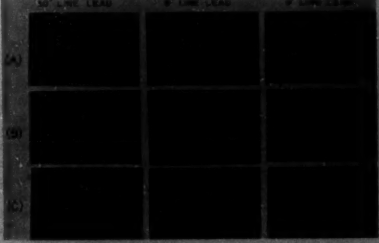


Fig. 11. The effect of applied wave shape on the response of typical dividers

A—Square-wave response
B—Triangular-wave response 0.19 microsecond to crest
C—Steep-front wave response 0.19 microsecond to crest

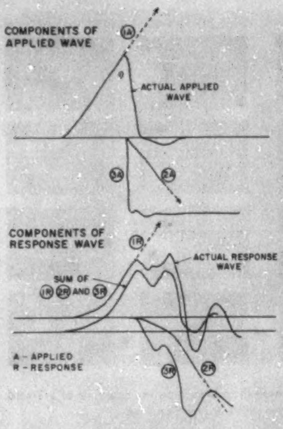


Fig. 12. Applied and response wave components of oscillogram 558

where

V_{kv} = steep front rating of divider and
$c=\dfrac{C}{V_{kv}}$ in farads per kilovolt
t = time in seconds
R = divider resistance in ohms

Thus the error that will exist in a divider of minimum length will be proportional to the rate of change of voltage of the applied wave. In the commercial impulse testing of transformers the steep-front wave for the higher voltage classes has a rate of rise of 1,000 kv/μsec. For such tests the error will be

$$\text{Per cent error} = 670R(10^{-6}) \qquad (9)$$

Assuming that 3% error can be tolerated the maximum value of R is equal to 4,500 ohms. This means that, using a divider of minimum length for the voltage being measured, the resistance should not be over 4,500 ohms if the error is to be kept to 3% for measurement of 1,000-kv/μsec steep-front waves. If the divider is not of minimum length consistent with the voltage to be measured, or if the resistance is greater than 4,500 ohms, the error will be greater than 3%. The present AIEE recommendation[7] considers only the time to crest and does not bring voltage magnitude, which involves divider height or capacitance, into the picture. For a 1,000-kv/μsec steep-front wave cresting at 1.2 μsec a divider of 12,000 ohms is permitted. This would produce an error of 8% compared to the AIEE claim of 5%.

Waves with higher rates of rise can be measured by reducing the resistance. For example, a divider of 1,000 ohms should be capable of measuring waves with rates of rise of 4,500 kv/μsec. However, as the resistance is decreased a value will be reached where the divider will develop large oscillations and cause considerable distortion of the response wave. Inspection of Figs. 14 and 15 shows that such a condition exists at less than 0.2 μsec for the 28-card, 1,680-ohm divider and less than 0.1 μsec for the 7-card 1,050-ohm divider. Assigning to these dividers a steep-front voltage rating of 75 kv per card, the rate of rise at which serious oscillations develop is about 10,000 kv/μsec for the 28-card divider and about 5,000 kv/μsec for the 7-card divider. The reason that the 7-card divider has a lower maximum rate of rise than the 28-card divider is at least partly due to the fact that the inductance per card of the small divider was greater than for the large divider. At any rate for the range of dividers tested, having steep-front voltage ratings from 525 to 2,100 kv and resistance values approaching a lower limit of 1,000 ohms, rates of rise of 5,000 kv/μsec do not cause serious oscillations within the divider. Thus within these limits, equation 7 can be used. That is:

$$\text{Per cent Error} = 0.67R\frac{V_{kv}}{t}10^{-12} \qquad (10)$$

when $\dfrac{V_{kv}}{t} \leq 5{,}000(10^{6})$

and $R \geq 1{,}000$

To summarize, vertical resistance dividers of the type used in this investigation will be accurate to within 3% if the resistance is not greater than shown in Table II and if a divider of minimum length is used.

The oscillations developed within the divider imposed the limitation of 4,500 kv/μsec. In order to measure faster rates of rise with less than 3% error it would be necessary to reduce either the divider resistance or stray capacitance. The capacitance can be reduced only slightly since even if the resistance cards were decreased in cross section to that of a no. 16 wire the capacitance would be decreased by less than 50%. The only alternative then for unshielded resistance dividers appears to be to reduce the divider resistance. In order to do this the inductance must be reduced, otherwise serious internal oscillations will develop. If the inductance per card is reduced by a factor of 10, as has been accomplished by Park, the over-all resistance may be decreased by a factor of three while still maintaining the same degree of damping. This would then allow a threefold increase in the maximum rate of rise which could be measured with no greater error than the present system.

All factors considered, it seems probable that uncompensated vertical resistance dividers can be further improved to the point where they can be used to measure with good accuracy rates of rise of the order of 10,000 kv/μsec. Beyond this point it may be necessary to provide some means of compensation or to develop new techniques. The most common approach for providing compensation is through the use of a line shield physically surrounding the upper portion of the divider. Ideally such a shield would be proportioned so the capacitive currents from the shield to any point on the divider would exactly equal the capacitive current from that point to ground. The proper shield is generally determined by a trial and error process aided by flux plotting. The major drawback to the use of a line shield is that it causes an increase in the capacitance to ground at the line end of the divider which tends to aggravate the oscillations involving the line lead. These oscillations make the voltage at the

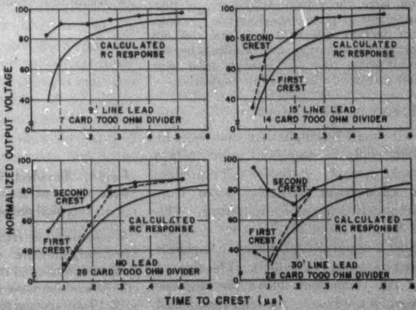


Fig. 13. The effect of physical height and line-lead length on the normalized voltage response of 7,000-ohm dividers. Steep-front waves applied

divider different from the voltage at the test piece end of the line lead. The increased oscillations may then offset the improved response of the divider proper.

Use of Square-Wave Response to Calculate Response to Other Waves

The square-wave response of a divider provides a general description of the performance of the divider, much the same as the complete frequency response characteristics do. A simple means of relating the square-wave response and the response to any applied wave will be developed.

The following symbols will be used:

U = amplitude of applied square wave = 1.0
U_R = normalized amplitude of response to applied square wave = 1.0 after steady state conditions have been reached
$Y(t)$ = any applied wave
Y_n = amplitude of applied wave at time t_n
$\Delta Y_n = Y_n - Y_{n-1}$ = incremental increase in amplitude during the time interval Δt_n
$Y_R(t)$ = response to $Y(t)$
Y_{Rn} = amplitude of response wave at time t_n

It can be shown that any applied wave can be visualized as a series of square waves as shown in Fig. 16. Assuming equal time intervals are chosen the amplitude of each square wave is proportional to the slope of the applied wave $Y(t)$ at the time that the square-wave component starts. The amplitude of the applied wave at any time (t) would then be equal to the sum of the amplitudes of the square-wave components. For instance:

at t_3,

$$Y_3 = U_A + U_B + U_C$$
$$= U\Delta Y_1 + U\Delta Y_2 + U\Delta Y_3$$

at t_n,

$$Y_n \sum_{s=1}^{s=n} U\Delta Y_s$$

In the limit as $\Delta Y \rightarrow 0$ this process becomes the equivalent of Duhamel's Theorem.

In a like manner the response to $Y(t)$ would be composed of the sum of the responses to the successively applied square waves with proper time relations observed. For example,

at t_3,

$$Y_{R3} = U_{R3}(\Delta Y_1) + U_{R2}(\Delta Y_2) + U_{R1}(\Delta Y_3)$$

at t_n,

$$Y_{Rn} = \sum_{s=1}^{s=n} U_{R(n+1-s)}\Delta(Y_s)$$

An example of a response wave calculated on the basis of the square-wave response is shown in Fig. 17. The square- and steep-front applied and response waves were taken from Fig. 8, oscillograms 574 and 575. The response waves have been normalized so that an accuracy comparison may be made directly. The applied steep-front wave was divided into about 40 equal time intervals over the range 0 to 0.8 μsec. This was the equivalent of breaking the applied wave into 40 square-wave components. The summation calculations were performed with the aid of a desk-type calculator in approximately an hour. It can be seen that the calculated and actual response waves agree very well.

As a method of checking dividers the square-wave approach has several advantages. First, the response to any applied wave may be determined easily and quickly. The arithmetical computations are readily adaptable to electronic data-processing methods. Second the square-wave response itself may be determined readily with equipment which is commercially available. Finally, the response characteristics can be determined without comparing applied and response waves directly. It is not necessary to know the exact gain of the amplifier; it is only necessary to know that its rise time is much faster than the rise time of the divider and that its gain thereafter is constant. All calculations are based on the steady-state amplitude of the square wave response. If the performance of the amplifier is unknown its gain characteristics may be determined by use of the techniques described earlier in this paper.

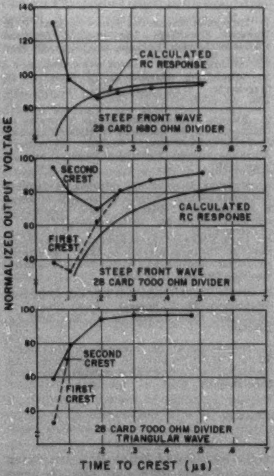


Fig. 14. The effect of resistance and wave shape on the normalized voltage response of a 28-card divider with a 30-foot line lead

Table II. Recommended Divider Resistance Values

Rate of Rise of Applied Wave, Kv/μsec	Divider Total Resistance, Ohms
1,000	4,500
2,000	2,250
3,000	1,500
4,000	1,125
4,500	1,000

Conclusions

1. Two methods for determining the response of impulse voltage dividers have been developed. One of these involves a comparison of the divider input and response waves through the use of a suitable amplifier, and includes a calibration of the test equipment. The second involves a calculation of the response to any applied wave based on the measured response to a square wave.

2. When the divider is of minimum length, the divider elements used by the authors have a capacitance of about 0.04 μμf/kv and an inductance of about 0.04 microhenries/kv.

3. With short line lead, resistance greater than 1,000 ohms, and rate of volt-

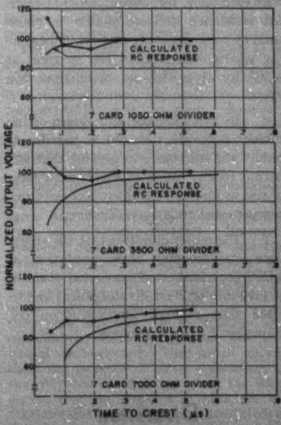
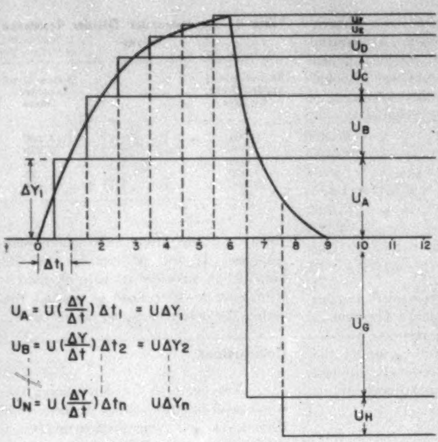


Fig. 15. The effect of resistance on the normalized response of a 7-card divider with a 9-foot line lead. Steep-front waves applied



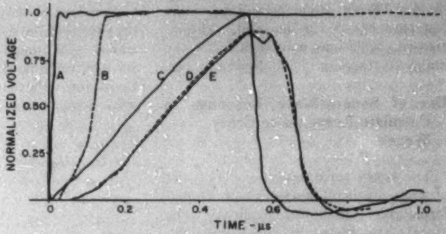


Fig. 16 (left). Representation of an impulse wave by a series of square waves

Fig. 17 (above). Results of calculated response based on square-wave response

A—Applied square wave
B—Response to A
C—Applied steep-front wave
D—Calculated response to C
E—Actual response to C

age rise less than 4,500 kv/μsec, the error in measuring steep-front waves with a divider composed of such elements can be readily calculated.

4. To limit the error to 3% when measuring a steep-front wave with a rate of rise of 1,000 kv/μsec the resistance of such a divider should not exceed 4,500 ohms.

5. Oscillations, either within the divider or involving the line lead, will limit the short-time response of resistance dividers. When they have been handled with care, however, uncompensated vertical resistance dividers similar to those used by the authors are capable of measuring waves with rates of rise as great as 4,500 kv/μsec with no more than 3% error.

6. By reducing the inductance of the divider elements it should be possible to reduce the resistance to a value that will permit measurement of waves with rates of rise as fast as 10,000 kv/μsec with uncompensated resistance dividers.

7. The paper indicates that the short-time measurement recommendation in AIEE Standard no. 4 may need revision to take into account divider height and to place more realistic limitations on the allowable divider resistance.

Appendix

Griscom, Lloyd, and Hileman[4] have derived the following equation for the response of resistance dividers assuming a constant stray capacitance per foot of length. The applied wave has a linear rising front of slope m.

$$E_R = m\left[t - \frac{2T}{\pi^2}(0.817) + \frac{2T}{\pi^2}\times\left(\epsilon^{-\frac{\pi^2 t}{T}} - \frac{1}{4}\epsilon^{-\frac{4\pi^2 t}{T}} + \frac{1}{9}\epsilon^{-\frac{9\pi^2 t}{T}}\ldots\right)\right] \quad (11)$$

where E_R = per unit voltage across the low-voltage resistor at bottom end of the divider.

This can be written as

$$E_R = 1 - \frac{2RC}{\pi^2 t}(0.817) + \frac{2RC}{\pi^2 t}\times\left(\epsilon^{-\frac{\pi^2 t}{RC}} - \frac{1}{4}\epsilon^{-\frac{4\pi^2 t}{RC}} + \frac{1}{9}\epsilon^{-\frac{9\pi^2 t}{RC}}\ldots\right) \quad (12)$$

where

R = total resistance
C = total stray capacitance

The error in per cent will be

$$\text{Error} = \frac{1-E_R}{1}100 = \frac{2RC}{\pi^2 t}(0.817) + \frac{2RC}{\pi^2 t}\times\left(\epsilon^{-\frac{\pi^2 t}{RC}} - \frac{1}{4}\epsilon^{-\frac{4\pi^2 t}{RC}} + \frac{1}{9}\epsilon^{-\frac{9\pi^2 t}{RC}}\ldots\right)100 \quad (13)$$

When $\frac{\pi^2 t}{RC} \geq 5$, that is $RC \leq 2t$, all terms but the first are essentially zero and

$$\text{Error} = \frac{RC}{6t}100 \quad (14)$$

When $RC = 2t$, the error is 33%. For smaller values of RC the error will be correspondingly less.

References

1. Errors in Recording Surge Voltage, P. R. Howard. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. II, Aug. 1952, pp. 371–83.

2. Short-Time Spark-Over of Gaps, J. H. Hagenguth. *AIEE Transactions*, vol. 56, Jan. 1937, pp. 67–76.

3. Errors in Measuring Surge Voltages by Oscillograph, M. Bochman, N. Hyltén-Cavallius. *Technical Achievements*, Research Laboratory, Allmänna Svenska Elektriska, Vasteras, Sweden, Mar. 1946.

4. Voltage Divider for Measuring Impulse Voltages on Transmission Lines, S. B. Griscom, B. L. Lloyd, A. R. Hileman. *AIEE Transactions*, vol. 73, pt. III-A, Apr. 1954, pp. 228–37.

5. Peak-Voltage Measurements of Standard Impulse Voltage Waves, A. Aked. *Monograph no. 153M*, Institution of Electrical Engineers, London, England, Nov. 1955.

6. Etude des Diviseurs de Tension en Regime Permanent, M. A. Romano. *Revue Générale de l'Électricité*, Paris, France, vol. 65, May 1956, pp. 289–300.

7. Measurement of Voltage in Dielectric Tests. *AIEE Standard no. 4*, Sept. 1953; *ASA Standard C 68.1*, American Standards Association, New York, N. Y.

◆

Discussion

S. B. Griscom, B. L. Lloyd, and A. R. Hileman (Westinghouse Electric Corporation, East Pittsburgh, Pa.): The authors have chosen a very important subject for investigation and one deserving the concentrated attention of all those concerned with high-voltage impulse measurements. The paper presents an interesting application of low-voltage pulse techniques to the verification of the accuracy of high-voltage measuring equipment, and the results and experience cited should be valuable to other investigators in this field. Unfortunately, we find that we are unable to agree with the authors' recommended values of divider

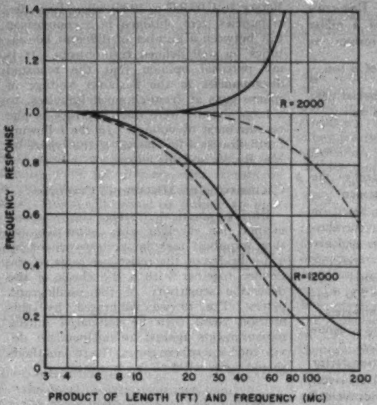


Fig. 18. Frequency response curves of resistance dividers, neglecting the resistance change with frequency

Stray capacitance = 5 $\mu\mu f$ per foot
Series inductance = 5 microhenrys per foot
———————— Includes series inductance and stray capacitance
— — — — Includes stray capacitance only

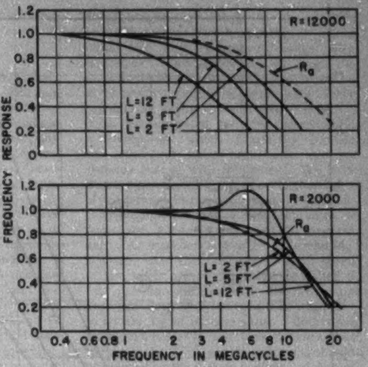


Fig. 19. Frequency response curves for resistance dividers

Stray capacitance = 5 $\mu\mu f$ per foot
Series inductance = 5 microhenrys per foot
Resistance change with frequency, see Fig. 3 of reference 4 of the paper

resistance, Table II, for measuring the specified rates of rise with an error of less than 3%.

Before discussing this table, we would like to comment on the relative merits of frequency response versus pulse techniques for evaluating the performance of dividers. One must first define whether the evaluation desired is the relative response of one divider compared to another or whether it is the absolute response of a divider to a specified wave form. For determination of relative response, which is particularly appropriate for the development of new divider systems, we feel that frequency response has many advantages over pulse techniques. Not only does it permit a ready comparison of the improvement accomplished, it greatly assists in segregating, analyzing, and correcting the various possible sources of error. For example, a resistance divider which yields underresponse on a front-chopped wave can do so because of the effect of stray capacitance, inductance or because of resistance variation with frequency. With tests or calculations involving pulse techniques, determination of the source of error is often difficult; using frequency response, it is much more obvious.

For evaluating the absolute performance of a divider to a specific wave, the use of frequency response requires establishing the frequency content of the wave, operating upon the components with the response characteristics, and recombining to obtain the resulting output wave. This appears to be somewhat cumbersome, and might make it seem that pulse techniques are better for determining absolute response. However, when pulses are employed, the response to that particular wave form is determined, but to no other form. All laboratory test circuits have somewhat different characteristics, and no two waves from different laboratories will have identical wave forms, particularly for front-chopped waves, even though the waves have the same nominal description. For example, the authors' wave shape shown in Fig. 12 is rounded at the crest and would appear to have a crest voltage of perhaps 5% higher if the chopped tail were extrapolated back up to the linearly rising front. Now if a divider gives 3% error with the authors' wave applied, the error would be nearer 8% if the theoretical chopped wave had been achieved. Consequently, for measuring, say, a 0.2-μsec front-chopped wave, a specified divider might or might not be adequate depending upon how sharply the wave is chopped. Fundamentally, the frequency response curves contain all the information to evaluate a wave of any form; response to a particular pulse form does not.

Further, for Standards work it is particularly important that evaluation of divider performance be reproducible at any laboratory. Frequency response is reproducible; pulse work in general would be much more difficult to reproduce.

The frequency response method of evaluating voltage divider performance consists of measuring the output voltage magnitude due to the application of a variable frequency steady-state voltage to the divider. As a criticism of the frequency-response method, it has been said that the relative phase angles of the various frequency components must also be measured to adequately define the divider performance. However, this presents no problem for the calculated curves; in fact, the calculated frequency response contains both phase shift and amplitude terms. For inductive and capacitive effects, the phase angle response begins to fall away at a lower frequency than does the magnitude. However, in general, the relation between phase angle and magnitude will not vary greatly for resistance dividers with differing characteristics. Consequently it is our opinion that the relative response of resistance dividers can be accurately judged by considering the magnitude alone. In our work on dividers (reported in reference 4 to the paper) we did not attempt to measure phase-angle response. However, the excellent checks which we obtained between calculated and measured magnitude response convinced us that no significant error could exist in the angle calculations.

These comments are not intended as a criticism of pulse techniques as used by the authors. Since ultimately the desire is to measure pulses, obviously the divider response to pulses is important. However, we do feel that the industry has not adequately exploited the advantages which frequency-response methods have to offer.

The authors of this paper indicate that the principal source of error in resistance-type dividers is stray capacitance. We agree with this statement for physically long, high-resistance dividers. However, resistance change with frequency and the series inductance are major influences for dividers having either short lengths or low resistances. The curves of Fig. 18 of this discussion (reference 4 of the paper) show the influence of inductance and capacitance. Resistance change with frequency is neglected. The dotted curves show the response if stray capacitance is the only source of error, while the solid curves give the response if both the inductance and capacitive effects are included. The close correlation between the curves for the 12,000-ohm divider indicates that the inductance does not play an important role and can be neglected. However, the stray capacitance and series inductance

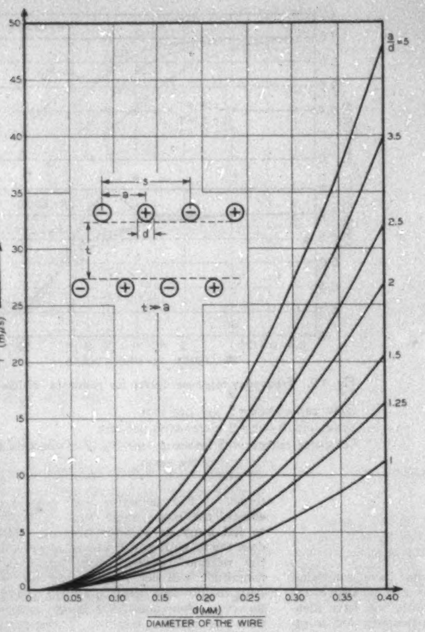


Fig. 20. Time-constant T of a bifilar wound resistor as function of d and $\frac{a}{d}$ for the special case $2a = s$

$T = \frac{L}{R} = \frac{100}{\rho} \quad kT' \text{m}\mu\text{s}$ time constant of the resistance

T'—value based on $\rho = 100$ microhm-centimeters and $k = 1$ (most resistors are wound with wire $\rho = 100$ microhm-centimeters)

ρ—microhm-centimeters

k—correction factor $k = 1.0 - 1.3$ (lowest for machines-wound resistance)

interact to produce severe over-response for the case of a 2,000-ohm divider. If capacitance is considered as the sole source of error, results would indicate that the divider is under-responding. The curves given in Fig. 19 of this discussion (reference 4 of the paper) show calculated frequency response curves for 12,000 and 2,000-ohm dividers with length as a parameter. The dotted curve marked R_a is the frequency response if the sole source of error is the resistance change with frequency. To determine the effect of each of the parameters, the curves of Figs. 18 and 19 should be compared.

First consider the case of the 2,000-ohm divider. The stray capacitance and inductance tend to give a condition of severe over-response. However, the effect of resistance change with frequency tends to give a condition of under-response. For small divider lengths, the resistance change with frequency becomes the principal source of error. For physically long dividers, the inductance and series capacitance increase the response into the area of over-response. However, the resistance change with frequency finally predominates and pulls the curve down.

Next consider the 12,000-ohm divider. For long dividers, the principal source of error is the stray capacitance. For shorter divider lengths, the resistance change with the frequency has a larger effect. However, the capacitive error still predominates.

Therefore, all factors affecting divider performance must be thoroughly evaluated and investigated to determine the source or sources of error. Since the table of recommended divider resistances in this paper is based upon stray capacitance, in the presence of constant resistance, as the only source of error, we cannot agree with the values tabulated.

As a final point, the authors suggest the way to overcome the effect of stray capacitance is to lower the total resistance. However, the resistance can easily be lowered to the point that errors due to the inductance acting with the stray capacitance will be much greater than the gain due to a lower RC product. The authors point out that grading shields can be added to compensate for the stray capacitance at the sacrifice of greatly increased capacitance loading at the divider line terminals. We suggest that a more satisfactory approach is to add small lumped capacitors across the resistor units as was done for the divider described in reference 4 of the paper. This approach eliminates the high capacitance terminal loading and eliminates complex design of elaborate shielding structures.

N. Hyltén-Cavallius (ASEA High Voltage Laboratory, Ludvika, Sweden): The authors are to be complimented for a thorough investigation and a valuable contribution in the important field of high-voltage measuring technique. The importance of good methods for checking impulse measuring circuits may be illustrated by the report by Berger to CIGRE in 1956[1] where large discrepancies were obtained in a comparison test between a number of different laboratories on well defined test objects. It is my personal opinion that the reported discrepancies in the flashover voltage of sphere gaps at front-chopped impulses, to a large extent, were due to errors in the measurement of voltage. In the following I will discuss a few points of the report by Mr. Rohlfs and his colleagues.

Calibration of Measuring Circuits

During practical testing the simplest method for checking the calibration (at slow impulses) may be a measurement by a d-c bridge of the resistance ratio of the divider together with a d-c check of the deflection sensitivity of the oscillograph plates. The device calibrated by this method could then be compared during measurements against an independant device such as a sphere gap. The two methods should agree within a few per cent.

Divider

I was very glad to see that the divider is defined as the whole measuring circuit from the test object to the oscillograph plates (including the leads connecting the divider to the test object). The dividers return lead on the earth side could, however, need a further study. In our laboratory we had a great many difficulties (disturbances) caused by circulating currents in this lead.

To avoid this, all the floors of our impulse testing plants have been covered with sheets of brass or aluminium. All objects, such as the impulse generator, the test object, the divider, and the oscillograph, are connected directly to these floors via minimum length of lead. The arrangement has resulted in an important reduction of disturbances.

Noninductivity Wound Resistors

I was slightly astonished at the high inductivity of the card resistors, especially those for lower resistor values, e.g., 20-ohm $L/R = 0.1$ μsec. Fig. 20 is a diagram for calculation of the inductive time constant of such a resistor. It has proved to agree well with measured values. By design, according to this diagram, it should generally be possible to reach time constants less than 0.01 μsec even in the low ohmic range.

Oscillations of Resistance Dividers

The authors refer these oscillations to the inductivity of the lead connecting the divider to the test object and the effective line-end capacity of the divider itself. It should, however, be observed that the lead itself oscillates as an approximately open "long line." The frequencies obtained by this agree fairly well with those found in Fig. 8.

They also refer to internal oscillations within the divider which could be reduced by a suitable choice of the divider resistance. According to personal experience I would, however, sooner refer these internal oscillations to the inductivity of the low-voltage arm of the divider. Such low-voltage arms are difficult to design and have in general much more inductivity than we want to believe. This inductivity "amplifies" the oscillations which for instance may be present in the lead connecting the divider to the test object.

ACCURACY OF VERTICAL RESISTANCE DIVIDERS

The authors give in Table II divider total resistances against rate of rise of applied wave for dividers with minimum length. From the standardizing point of view, however, it would be simpler to put this the other way round, especially as the expression "minimum length" may be subjected to much discussion. A rewriting of the authors' equation 5 gives the relative error Δ as:

$$\Delta=\frac{1}{6}\frac{RC}{t}10^{-6}\simeq 0.6\frac{(R)(l)}{t}10^{-6}$$

where

Δ = the relative error
R = the total divider resistance ohm
C = the total divider capacitance to earth ($C\simeq 3.6l$ $\mu\mu$f)
l = the divider length foot
t = time to chopping microsecond

If an error of 5% ($\Delta=0.05$) is permitted this gives

$$\frac{Rl}{t}\simeq 8(10^4)$$

An additional note would prevent the use of too low-ohmic dividers.

I should add a hope that the points where I have shown a slight disagreement with the authors do not conceal my admiration for their excellent work.

REFERENCES

1. COMPARATIVE TESTS ON SPARK-GAPS, K. Berger. *Report no. 326*, CIGRE, Paris, France, 1956.

2. BIFILAR WOUND RESISTANCES, E. Pirktl. *Internal Report Thl 1P-13, 10.9, 1955*, Allmänna Svenska Elketriska, Aktiebologet, Ludvika, Sweden

J. H. Park (National Bureau of Standards, Washington, D. C.): The authors have demonstrated the usefulness of applying low-voltage repeated pulse techniques for comparing the performance of high-voltage dividers. The over-all accuracy is not definitely stated. It appears from their oscillograms that their reference square-wave response time is about 0.02 μsec, which for waves cresting at 0.5 μsec would represent an error of 4%. This is, of course, the total uncertainty introduced by their pulse generator, attenuator, amplifier, and oscilloscope. It does not limit the use of this method to waves cresting at more than 0.5 μsec.

Recent advances in high frequency and repeated steep-rising pulse techniques have made such methods very useful in the high-voltage surge measurement field. However, it should be emphasized that the final proof of accuracy for steep-front high-voltage surge measurements must include a test at or near rated voltage. If sphere gap flashover voltages obtained using linear rising voltages of various rates of rise can be established by agreement among several laboratories, then such data could be used to check measurement methods in other laboratories.

A. F. Rohlfs, J. S. Kresge, and **F. A. Fisher:** Messrs. Griscom, Lloyd, and Hileman dwell at length on the relative merits of frequency response versus pulse techniques. They list two alternatives that a test may accomplish: a relative evaluation of several dividers, or the absolute response of a given divider to a specified wave form. They have apparently missed the significance of our use of the square-wave response to calculate the response of a divider to any applied wave. We have not only pointed out that the square-wave response provides a general description of the divider's pulse response, but we have given a detailed procedure for finding the response to any wave. Figs. 16 and 17 of the paper illustrate the method and demonstrate the accuracy of the technique.

In theory, of course, a Fourier analysis of a wave may be made, the frequency components may be acted upon by the frequency characteristics of the divider, and the resulting components may be recombined to provide the response wave. Unless the applied wave and the frequency response characteristics (including the phase characteristics) can be expressed as analytic functions, however, this is almost prohibitively difficult. Although accurate results might be obtained without considering phase characteristics, this should be rigorously proven before such a simplification is made. If Messrs. Griscom Lloyd, and Hileman have developed a practical procedure for relating the frequency response to wave-shape distortion and magnitude errors for impulse waves (even approximately) it would be a welcome and valuable addition to the literature. It is certainly desirable to have at least two independent methods for determining a divider's response to impulse waves.

As a method of comparing the relative performance of dividers the square-wave approach has advantages also, particularly in that the square-wave response may be determined much faster than the frequency response. We further feel that pulse techniques, particularly the square-wave technique which does not involve the pulse-shaping network of Fig. 2 are just as reproducible as frequency response techniques.

Measurements of the actual steep-front wave response, of course, do not determine the general response of a divider system. They do, however, have two very important advantages; they provide an absolute measure of the accuracy and they illustrate clearly the type of distortion a given divider produces on a wave.

We agree that an applied steep-front wave which rounds over near crest is measured with less error than one which does not round over. In fact, this is stated in the section on amplitude response as one of the reasons why the measured errors were less than the calculated errors. We feel that it is necessary to assume the worst condition in calculating errors in order to cover all possibilities and therefore used the idealized steep front wave (no round over) as the basis for calculation of error.

They object to Table II, the recommended resistance values, on the grounds that no account has been taken of series inductance and resistance change with frequency. This is not so. Although the table is derived from calculations based on a constant resistance, and with stray capacitance being the sole source of error, it is substantiated by the experimental measurements which inherently include the effects of series inductance and resistance change with frequency. We have devoted considerable attention to the effects of inductance in the paper and have specifically pointed out that a minimum value of resistance must be exceeded to prevent undue internal oscillations from developing. We feel that the resistance values listed in the table are conservative and that if the recommendations are followed any practical steep front wave will be measured to within 3% of its true value. Both their work (reference 4) and ours indicate that the present AIEE recommendations should be revised to take more explicit account of divider length and inductance. Any specific proposals would be very valuable.

We agree with Mr. Hylten-Cavallius that the primary calibration of a resistance divider should be done with a d-c bridge, but feel that the limitations on performance should be established and the divider used within its limitations. We have used one method of establishing these limitations; other methods are needed.

Grounding techniques certainly are important. A solid metal ground plane perhaps represents the best solution, although its superiority over a well designed counterpoise would have to be investigated. He expresses concern over the high L/R ratios for the low value resistors. These refer only to low-resistance high-voltage cards which are not generally used in dividers. The low voltage arm of the dividers were resistors of a different construction (as discussed in the paper) and have much better L/R characteristics because of their small physical size.

We cannot agree with him that the internal oscillations should be ascribed to the inductance of the low voltage arm. We have observed oscillations even when the output of the divider was fed directly into the coaxial deflection cable with no resistor at the input end and with a high quality terminating resistor at the output end. Oscillations within the high voltage arm should not be too surprising for it is, after all, an RLC network. Any such network can be excited into oscillation if the resistance is too low.

Mr. Hylten-Cavallius has misinterpreted our capacitance figure which works out as 4.5 $\mu\mu$f/foot of resistance winding. Each 10-inch card contains only 8 inches of winding and due to the method of mounting, contributes 8 inches to the over-all height of the divider. His equations then become as follows

$$\Delta=1/6\,\frac{RC}{t}10^{-6}\simeq 0.75\,\frac{Rl}{t}10^{-6}$$

and for an allowable error of 5%, $Rl/t=6.67(10^4)$ ohms. His other points are well taken. We introduced the rate of rise concept as a link with transformer steep-front wave testing. In general it is more straightforward to speak of the time to crest of the wave than to speak of the rate of rise. From a measurement viewpoint there is no lower limit on divider resistance as long as the divider inductance is such that internal oscillations do not develop.

Mr. Park measures a response time of 0.02 μsec for our reference square wave. This primarily reflects the rise time of the amplifier. The rise time of the square wave itself, as measured directly on the deflection plates of the cathode-ray oscillograph, is closer to 0.002 μsec. We feel that the over-all error in our measurements system is less than 3% for the types of applied waves considered here. Certainly a test at rated voltage would be desirable. Even at low voltages several different approaches would be of value in order to provide additional verification of the results.

Dynamic Operation of Magnetic Amplifiers with Real-Core and Rectifier Functions

H. C. BOURNE, JR.
MEMBER AIEE

D. NITZAN
STUDENT MEMBER AIEE

As magnetic-amplifier circuitry has evolved from the saturable-reactor circuits to the self-saturating types, the basic limitations on magnetic-amplifier performance are inevitably associated with the nonlinear materials of the cores and rectifiers. The analytic descriptions of the core and rectifier characteristics introduce the critical parameters in a representation of the dynamic operation of the self-saturating types in contrast to the dependence of the saturable-reactor circuits on essentially linear resistance functions.[1] Unrealistic assumptions concerning these nonlinear material functions, which under some circumstances might well be idealized, give results which are of limited usefulness to those concerned with the design and analysis of magnetic amplifiers as system components.

Many analyses have been developed which are unquestionably useful under certain restrictions. Magnetic amplifiers used essentially as power amplifiers or power modulators with a relatively small control-circuit impedance and idealized rectifiers have been thoroughly treated.[2,3] Magnetic amplifiers with idealized cores but with imperfect rectifiers and no limitations on the size of the control-circuit impedance have been analyzed by a comprehensive technique involving difference equations.[1] Other analyses[4-6] involve the half-cycle-response circuits which are not directly of concern in this paper.

Magnetic amplifiers are being used increasingly as voltage amplifiers in feedback control systems with the emphasis on voltage-gain bandwidth requirements rather than power-gain requirements. This use has pinpointed the need for a dynamic transfer function of the self-saturating magnetic amplifier which is generally valid without being prohibitively complicated. The purpose of this analysis is to derive a representation in terms of the published properties of magnetic cores and rectifiers which will accurately predict the performance of the device with a minimum number of restrictions.

The familiar a-c self-saturating magnetic-amplifier circuit is chosen as the example of an analysis which may be easily applied to the center-tap and bridge circuits. The techniques involving difference equations, used so effectively by Johannessen,[1] are adapted for use here. The difference equation by which the operation of all magnetic amplifiers is inherently described, is interpreted in terms of a transfer function of the s-variable (ϵ^{Ts}). Although the transfer function as a ratio of polynomials of the s-variable (Laplace) becomes a rough approximation for fast-response magnetic amplifiers, the approximation may be derived and used with good results for slower response times.

One of the powerful methods used to analyze nonlinear phenomena consists of dividing the phenomena into regions in which the behavior is piecewise-linear to a good approximation. The method is very suitable in the analysis of magnetic amplifiers since the hysteresis loops of modern core materials are approximately rectangular in shape. Rectifier characteristics are also amenable to this representation. In this case, operation is divided into four modes, depending on the state of each core, and the appropriate equations are developed for each mode. Boundary conditions are then matched to complete the solution.

To avoid an analysis so complex as to be only of theoretical interest, approximations must be made to eliminate certain transcendental relations. Unfortunately, the magnetic amplifier is inescapably a highly complicated nonlinear device. However, in every case only those approximations are made which may be justified from the viewpoint of practical test results.

Modes of Operation

Basic Assumptions

The magnetic-amplifier circuit to be analyzed is shown in Fig. 1. The core characteristics are shown in Fig. 2. Although I_{cm} and I_{cm}' may be identified with the widths of the dynamic and static loops, such an identification is not necessary at this point. Operation is assumed on the larger loop when both cores are excited and along the smaller loop when only one core is excited. The rectifiers are assumed capable of representation by a backward and forward resistance. The excitation in the gate circuit is sinusoidal and, in order to simplify the analysis, the control source is chosen to be a full-wave rectified sinusoidal voltage. The choice of a direct control voltage would add much to the complexity of the theoretical analysis without giving any important difference in the practical results. If the gate circuit were excited by an alternating square wave and the control by the corresponding direct voltage, even less complexity would arise but the results would be somewhat less practical.

Mode A Operation

Assume that at $t=0$ core 2 is at a flux level ϕ_m and that core 1 is at ϕ_s. Since

$$e_s = E_{sp}\sin\omega t = \frac{\pi}{2}E_s\sin\omega t$$

Paper 57-94, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957. Manuscript submitted October 29, 1956; made available for printing June 11, 1957.

D. Nitzan is a student and H. C. Bourne, Jr., is on the faculty at the University of California, Berkeley, Calif.

This research was supported by the Air Force Cambridge Research Center, Air Research and Development Command, under contract no. *AF19(604)-1813*.

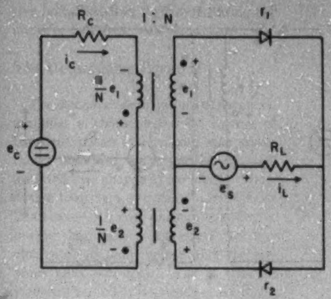


Fig. 1 (left). A-c self-saturated magnetic-amplifier circuit

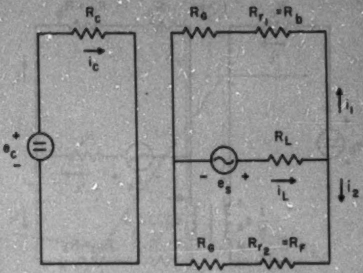


Fig. 3 (right). Equivalent circuit during mode A

and

$$e_c = E_{cp}\left|\sin \omega t\right| = \frac{\pi}{2} E_c \left|\sin \omega t\right|$$

both voltages are zero at $t=0$. As the positive half-cycle starts, both e_s and e_c increase, but still not enough to be able to supply the mmf (magnetomotive force) requirement of either of the two cores. At $\omega t=\beta_2$ core 2 becomes excited (point 7, Fig. 2). During the period $0<\omega t<\beta_2$ there is no coupling between the gate windings and the control windings. The equivalent circuit diagram is shown in Fig. 3, which defines mode A. It is assumed that rectifier r_1 is blocking and rectifier r_2 is unblocking the currents whose assumed positive directions are shown.

Examination of the circuit and recognition of the fact that $R_x >> R_w$, R_L, in which $R_x = R_g + R_b$ and $R_w = R_g + R_f$, show that core 2 becomes excited before core 1. The solution for the currents may be obtained and substituted into $Ni_2 + i_c = I_{cm}$, which describes the mmf requirements of core 2. The excitation angle of core 2 is then

$$\sin \beta_2 = \frac{(2/\pi) I_{cm}}{(NE_s/R_L+R_w)+E_c/R_c} \tag{1}$$

MODE B OPERATION

As soon as core 2 reaches excitation (point 7, Fig. 2), mode B starts. In this mode core 2 is excited, and its flux level rises from point 7 according to

$$\Delta\phi = (1/N_G)\int e_2 dt$$

whereas core 1 is still in saturation moving toward point 2. Three loop equations and the constraint imposed by the core, $Ni_2 + i_c = I_{cm}$, yield solutions for the currents and voltages of Fig. 4, the equivalent circuit of mode B. With the assumption that $R_x >> R_w$, R_L, then the exciting voltage across the core is given by

$$e_2 = \frac{N^2 R_c e_s + (R_w+R_L)(Ne_c - NR_c I_{cm})}{R_L+R_w+N^2 R_c} \tag{2}$$

Substitution of expressions for the currents into the constraint imposed by core 1, $Ni_1 - i_c = I_{cm}$, gives the boundary of mode B as

$$\sin \beta_1 = \frac{2}{\pi}\,\frac{I_{cm}}{N}\,\frac{2R_L + (2R_w + N^2 R_c)\dfrac{R_x}{R_x + N^2 R_c}}{E_s - NE_c\,\dfrac{R_x}{R_x+N^2R_c}} \tag{3}$$

Throughout the analysis of modes A and B and the modes that follow, the rectifiers are assumed blocking or unblocking as a condition of the mode. The actual existence of the mode and the fulfillment of the corresponding rectifier states are verified from the current solutions.

The analysis of modes A and B is based on a B-H loop with right-angled corners, a condition obviously not fulfilled with practical materials. Fortunately, since modes A and B will have a negligibly small effect on the dynamic representation, no error is introduced. The analysis of these modes is included for completeness and to justify later approximations.

MODE C OPERATION

When core 1 becomes excited (point 2, Fig. 2), mode C starts. The flux of core 2 continues to increase (toward point 8, Fig. 2) and the flux of core 1 begins to decrease (from point 2, Fig. 2, to point 3). The equivalent circuit during mode C is shown in Fig. 5. The solution to three loop equations and two core constraints yield these solutions for voltages and currents during mode C for $R_x >> R_w$, R_L:

$$i_1 = \frac{Ne_c + N^2 R_c I_{cm}/N}{R_x + N^2 R_c} \tag{4}$$

$$i_2 = \frac{-Ne_c + (N^2 R_c + 2R_x) I_{cm}/N}{R_x + N^2 R_c} \tag{5}$$

$$i_L = 2I_{cm}/N \tag{6}$$

$$i_c/N = \frac{Ne_c - R_x I_{cm}/N}{R_x + N^2 R_c} \tag{7}$$

$$e_1 = e_s - [Ne_c + (2R_L + 2R_w + N^2 R_c) I_{cm}/N]\,\frac{R_x}{R_x + N^2 R_c} \tag{8}$$

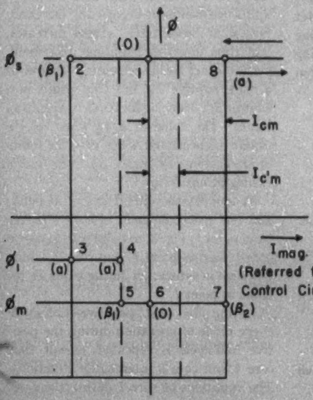


Fig. 2. Rectangular approximations of the core characteristics showing assumed flux variations during a positive half-cycle of the supply frequency

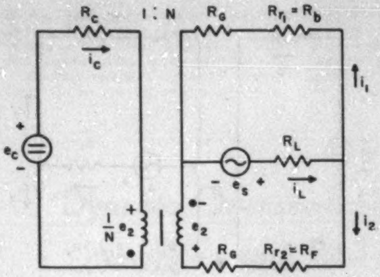


Fig. 4 (left). Equivalent circuit during mode B

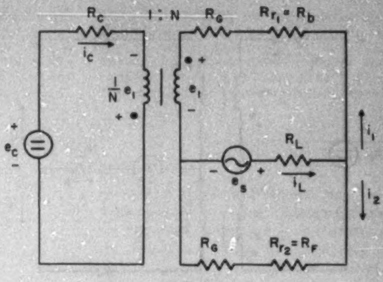


Fig. 6 (right). Equivalent circuit during mode D

$$e_2=e_s-\left(2R_L+2R_w-\frac{R_wN^2R_c}{R_x+N^2R_c}\right)I_{cm}/N \quad (9)$$

If the actual cores are replaced by ideal cores and current sources of value I_{cm}/N, these equations may be written using superposition without a formal solution of the loop or node and core-constraint equations.

During mode C the core fluxes are changing corresponding to the voltages e_1 and e_2. At $\omega t=\alpha$ the operating point of core 2 reaches saturation (point *8*, Fig. 2), and mode C ends. The flux of core 1 has decreased to ϕ_1 (point *3*, Fig. 2), which corresponds to a smaller flux change than has occurred in core 2 since $e_1<e_2$ from equations 8 and 9.

Mode D Operation

During mode D core 2 cannot absorb any voltage, since the saturation region is assumed to be horizontal and therefore $d\phi_2/dt=0$. The equivalent circuit is shown in Fig. 6 and the equilibrium equations may be written. The core constraint is now written $Ni_1-i_c=I_{cm}'$, which permits the flux excursion of core 1 to take place on a different path than that traveled in mode C. As previously mentioned, this new value of magnetizing current is often approximated with the static loop. The following equations during mode D are obtained with $R_x>>R_L$, R_w.

$$i_1=\frac{e_sR_w/(R_w+R_L)+Ne_c+NI_{cm}'R_c}{R_x+N^2R_c} \quad (10)$$

$$i_2=e_s/(R_w+R_L)-\frac{(Ne_c+NI_{cm}'R_c)R_L/(R_L+R_w)}{R_x+N^2R_c} \quad (11)$$

$$i_L=e_s/(R_w+R_L)+\frac{(Ne_c+NI_{cm}'R_c)R_w/(R_L+R_w)}{R_x+N^2R_c} \quad (12)$$

$$i_c/N=\frac{e_sR_w/(R_L+R_w)+Ne_c-R_xI_{cm}'/N}{R_x+N^2R_c} \quad (13)$$

$$e_1=e_s\frac{R_w}{R_L+R_w}\,\frac{N^2R_c}{R_x+N^2R_c}-(Ne_c+NI_{cm}'R_c)\frac{R_x}{R_x+N^2R_c} \quad (14)$$

Again, these equations may be written directly from the equivalent circuit by superposition.

The end of mode D occurs when e_s and e_c are so small that the total mmf of core 1 is too small to fulfill the mmf requirements of I_{cm}'. If the instant is denoted by $\omega t=\beta_1'$, then after $\omega t=\beta_1'$ cores 1 and 2 enter mode A, since in both cores the coupling between the control and the gate winding is zero. Mode D, therefore, extends from α to β_1' and is followed by mode A, which extends from β_1' to $\pi+\beta_2$. To find an expression for the boundary of mode D, the values of i_c and i_1 are substituted into the core constraint, $Ni_1-i_c=I_{cm}'$, to yield

$$\sin\beta_1'=\frac{2I_{cm}'/\pi}{NE_sR_w/R_x(R_L+R_w)-E_c/R_c} \quad (15)$$

An alternative procedure would be to equate e_1 to zero in equation 14.

In practical circuits the value of β_1' is not always close to π. With reference to Fig. 2, during mode D the operating point of core 1 drops from point *3* to point *5* (the transition from point *3* to point *4* is instantaneous and happens at $\omega t=\alpha$). The flux change during mode D is

$$\Delta\phi_D=\frac{1}{N_G\omega}\int_{\alpha}^{\beta_1'}e_{1D}d\omega t \quad (16)$$

where e_{1D} is the voltage appearing across core 1 during mode D, as given by equation 14. Aside from the fact that e_{1D} is small compared with e_{1C}, the value of $\Delta\phi_D$ becomes even smaller compared with $\Delta\phi_C$ due to limits of integration α and β_1'. As E_C becomes more negative β_1' increases, but, as is shown later, α also increases, and thus tends to maintain $\Delta\phi_D$ small. On the other hand, decreasing α requires a more positive E_C, but this results in a smaller β_1', which again tends to keep $\Delta\phi_D$ small. The value of $\Delta\phi_D$ when $\alpha<<\pi$ cannot be neglected, since now the limits of integration of e_C make $\Delta\phi_C$ comparable in magnitude to $\Delta\phi_D$.

At $\omega t=\beta_1'$, the circuit enters mode A, and the operating point of core 1 moves from point *5*, Fig. 2, to point *6*. At $\omega t=\pi$ core 1 reaches point *6*, and core 2, which has been in saturation during mode D, returns to point *1*.

During the next negative half-cycle every mode of operation during the positive half-cycle is repeated, except that core 1 and core 2 interchange functions. The equations of core 1 during the posi-

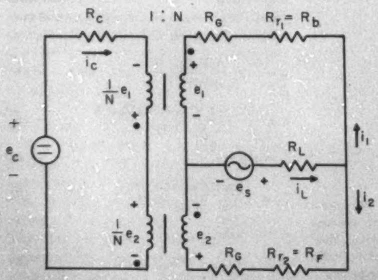


Fig. 5. Equivalent circuit during mode C

tive half-cycle hold for core 2 during the negative half-cycle, and vice versa.

The Determining Equation

The dynamic operation of the magnetic amplifier is characterized by a series of transient pulses of changing area essentially described by equation 12 and initiated by a change in the input. However a basic fact of operation is that the flux change of a core during the gating or output half-cycle must equal the flux change of the same core during the previous reset half-cycle. This is true for transient as well as for steady-state conditions. A flux difference equation based on this phenomenon is called the determining equation. Steady state is easily derived by setting each variable of the nth half-cycle equal to the corresponding one in the $(n+1)$th half-cycle. In terms of the flux changes in the gating and reset half-cycles, the determining equation is

$$\Sigma\Delta\phi_{rhc}(n)=\Sigma\Delta\phi_{ghc}(n+1) \tag{17}$$

$$\Delta\phi_{rhc_C}(n)+\Delta\phi_{rhc_D}(n)=\Delta\phi_{ghc_B}(n+1)+\Delta\phi_{ghc_C}(n+1) \tag{17A}$$

The flux changes of equations 17 and 17(A) may be expressed in terms of the corresponding volt-time integrals:

$$\frac{1}{\omega N_G}\int_{\beta_1(n)}^{\alpha(n)} e_{1C}(n)d\omega t+\frac{1}{\omega N_G}\int_{\alpha(n)}^{\beta_1'(n)} e_{1D}(n)d\omega t = \frac{1}{\omega N_G}\int_{\beta_2(n+1)}^{\beta_1(n+1)} e_{2B}(n+1)d\omega t+ \frac{1}{\omega N_G}\int_{\beta_1(n+1)}^{\alpha(n+1)} e_{2C}(n+1)d\omega t \tag{18}$$

in which e_{2B} is given by equation 2, e_{2C} by equation 9, e_{1C} by equation 8, and e_{1D} by equation 14. All quantities are indicated as functions of e_c, the control voltage. Substitution of the appropriate values for the voltages and the mode boundaries β_1, β_2, and β_1', and integrating, the firing angles $\alpha(n)$ and $\alpha(n+1)$ are obtained as functions of the control voltage in the nth and $(n+1)$th half-cycles, $E_c(n)$ and $E_c(n+1)$. The output voltage in the nth and the $(n+1)$th half-cycles may be obtained as the sum of the various contributions during modes A through D and consequently as functions of $E_c(n)$ and $E_c(n+1)$ In theory at least, the three equations

$$f_1[\alpha(n),\alpha(n+1)]=f_2[E_c(n),\ E_c(n+1)] \tag{18A}$$

$$E_L(n)=f_3[E_c(n),\ \alpha(n)] \tag{19}$$

$$E_L(n+1)=f_3[E_c(n+1),\ \alpha(n+1)] \tag{20}$$

may be solved to obtain the output difference equation

$$E_L(n+1)=f_4[E_c(n+1),\ E_c(n),\ E_L(n)] \tag{21}$$

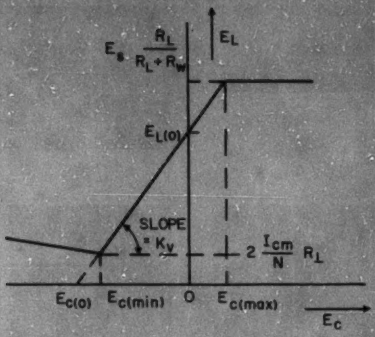


Fig. 7. Approximate linear transfer characteristic

In practice, such a solution would require a computer and the nonlinear difference equation resulting would be too complicated to be useful. These equations are treated in more detail elsewhere.[7,8]

Practical Output Difference Equation

Equations 18 through 21 become manageable if the effect of modes A and B are neglected in the formulation of the determining and output difference equations, i.e., $\beta_1\cong\beta_2\cong0$ and $\beta_1'\cong\pi$. These approximations may be justified from the analytic expressions and from experimental results. Under these conditions, equations 18(A), 19, and 20 become

$$\frac{1-\cos\alpha(n)}{2}\left[E_s-NE_c(n)\frac{R_x}{R_x+N^2R_c}\right]-\frac{\alpha(n)}{\pi}\frac{I_{cm}}{N}(2R_L+2R_w+N^2R_c)\times\frac{R_x}{R_x+N^2R_c}+\frac{1+\cos\alpha(n)}{2}\times\frac{E_sN^2R_c\dfrac{R_w}{R_w+R_L}-NE_c(n)R_x}{R_x+N^2R_c}-\frac{\pi-\alpha(n)}{\pi}I_{cm}'NR_c\frac{R_x}{R_x+N^2R_c}=\frac{1-\cos\alpha(n+1)}{2}E_s-\frac{\alpha(n+1)}{\pi}\frac{I_{cm}}{N}\times\left[2(R_L+R_w)-\frac{R_wN^2R_c}{R_x+N^2R_c}\right] \tag{18B}$$

$$E_L(n)=\frac{1+\cos\alpha(n)}{2}E_s\frac{R_L}{R_L+R_w}+2\frac{I_{cm}}{N}R_L\frac{\alpha(n)}{\pi} \tag{19A}$$

$$E_L(n+1)=\frac{1+\cos\alpha(n+1)}{2}\times E_s\frac{R_L}{R_L+R_w}+2\frac{I_{cm}}{N}R_L\frac{\alpha(n+1)}{\pi} \tag{20A}$$

Although these equations are still too formidable to use in obtaining a dynamic transfer function directly, they may be used to obtain practical and accurate results for the steady-state transfer characteristic.

In the steady state, equation 18(B) becomes

$$\frac{1+\cos\alpha}{2}=\frac{R_x}{NR_c}\times\frac{E_c+\left(\dfrac{\alpha}{\pi}I_{cm}+\dfrac{\pi-\alpha}{\pi}I_{cm}'\right)R_c}{E_s\dfrac{R_w}{R_L+R_w}} \tag{18C}$$

Equation 18(B) together with equation 19(A), which is identical to equation 20(A) in the steady state, may also be used as check points with results previously obtained in the literature. If $R_x\rightarrow\infty$ and $I_{cm}'\rightarrow0$, then

$$E_c=-\frac{\alpha}{\pi}I_{cm}R_c \tag{18D}$$

$$E_L=2\frac{I_{cm}}{N}R_L\frac{\alpha}{\pi}+\frac{1+\cos\alpha}{2}E_s\frac{R_L}{R_L+R_w} \tag{19B}$$

which are well-known expressions.[3] If the magnetizing currents are allowed to approach zero another familiar result is obtained.[1]

As the next step in obtaining a practical output difference equation, the contribution of the magnetizing current to the load current is neglected. This step is equivalent to discarding terms in equations 18(B), 19(A), and 20(A) comparable to $I_{cm}R_L$ or $I_{cm}R_w$ but not $I_{cm}R_c$ or $I_{cm}R_x$. With this approximation the output in the $(n+1)$th half-cycle may be obtained.

$$E_L(n+1)=M_1E_c(n)+M_2E_L(n)+M_1R_c\times\left[\frac{\alpha(n)}{\pi}(I_{cm}-I_{cm}')+I_{cm}'\right] \tag{21A}$$

In this equation the terms are as follow.

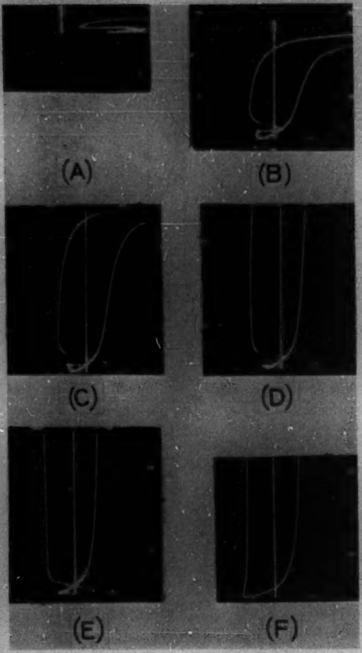


Fig. 8. Behavior of the resetting core on the B-H loop during mode D of operations for various magnetic amplifier outputs

A—Full-per-cent output
B—87½% output
C—75% output
D—50% output
E—25% output
F—Minimum output

$$M_1 = \frac{R_L}{R_L+R_w}\,\frac{NR_x}{R_x+N^2R_c}$$

$$M_2 = 1 - \frac{R_w}{R_L+R_w}\,\frac{N^2R_c}{R_x+N^2R_c}$$

$$\alpha(n) = \cos^{-1}\left[E_L(n)\,\frac{2}{E_s}\,\frac{R_L+R_w}{R_L} - 1\right]$$

The first three terms of equation 21(A) represent a result previously obtained.[1]

Equation 21(A) is a nonlinear difference equation and is still not directly applicable to the formulation of the dynamic transfer function. Two possible approximations are examined.

Approximation $\cos^{-1} x \cong (1-x)\pi/2$. This approximation corresponds to replacing the inverse cosine by a straight line through $x=0$ with the same slope as the original function. If the inverse cosine term in equation 21(A) is replaced by this approximation,

$$E_L(n+1) = K_1E_c(n) + K_2E_L(n) + K_3 \qquad (21\text{B})$$

where

$$K_1 = M_1 = \frac{R_L}{R_L+R_w}\,\frac{NR_x}{R_x+N^2R_c}$$

$$K_2 = 1 - \frac{N^2R_c}{R_x+N^2R_c}\left[\frac{R_w}{R_L+R_w} + \frac{2R_x(I_{cm}-I_{cm}')}{\pi NE_s}\right]$$

$$K_3 = \frac{R_L}{R_L+R_w}\,\frac{R_x}{R_x+N^2R_c} \times NR_c\left(\frac{\pi+2}{2\pi}\,I_{cm} + \frac{\pi-2}{2\pi}\,I_{cm}'\right)$$

Equation 21(B) is the output difference equation in a useful form.

Approximation $\cos^{-1} x \cong (1-x)\pi/2$. This approximation corresponds to replacing the inverse cosine by a straight line through $x=1$ and $x=-1$ and consequently gives a result indicative of the average slope. If the inverse cosine term in equation 21(A) is replaced by this approximation,

$$E_L(n+1) = L_1E_c(n) + L_2E_L(n) + L_3 \qquad (21\text{C})$$

where

$$L_1 = K_1 = M_1 = \frac{R_L}{R_L+R_w}\,\frac{NR_x}{R_x+N^2R_c}$$

$$L_2 = M_2 - \frac{M_1R_c(I_{cm}-I_{cm}')}{E_s\,\dfrac{R_L}{R_L+R_w}} = 1 - \frac{N^2R_c}{R_x+N^2R_c}\left[\frac{R_w}{R_L+R_w} + \frac{R_x(I_{cm}-I_{cm}')}{NE_s}\right]$$

$$L_3 = \frac{R_L}{R_L+R_w}\,\frac{R_x}{R_x+N^2R_c}\,NR_cI_{cm}$$

In the next section equation 21(B) is examined in some detail. However, by the very simple expedient of replacing $(\pi/2)E_s$ by E_s, the results will apply directly to equation 21(C), since L_2 differs from K_2 by this factor. The minor difference between K_3 and L_3 is insignificant and represents a linear shift of the transfer characteristic having no effect on gain or time response.

Dynamic Transfer Function

LINEAR STEADY-STATE TRANSFER CHARACTERISTIC

A simple method to predict an approximate steady-state transfer characteristic of the magnetic amplifier is sometimes important in the design procedure of a servo system. In the steady-state condition, $E_L(n+1) = E_L(n) = E_L$ and $E_c(n) = E_c$. Equation 21(B) becomes

$$E_L = \frac{K_1}{1-K_2}\,E_c + \frac{K_3}{1-K_2} \qquad (22)$$

Substituting the values of K_1, K_2, and

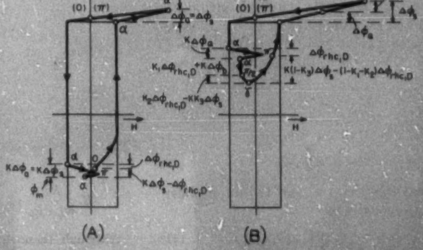

B
(0)
(π)
α
α
$\Delta\phi_a = \Delta\phi_s$
H
α
0
α
π
$K\Delta\phi_a = K\Delta\phi_s$
ϕ_m
$\Delta\phi_{rhc,D}$
$K\Delta\phi_s - \Delta\phi_{rhc,D}$
(A)
B
(0)
(π)
π/2
$\Delta\phi_b$
$\Delta\phi_s$
$\Delta\phi_a$
$K\Delta\phi_a$
α
π
α
π/2
γ
$\Delta\phi_{rhc,D}$
$K_1\Delta\phi_{rhc,D} + K\Delta\phi_b$
$K(1-K_3)\Delta\phi_s - (1-K_1-K_2)\Delta\phi_{rhc,D}$
$K_2\Delta\phi_{rhc,D} - KK_3\Delta\phi_s$
H
(B)

Fig. 9. Effect of saturation inductance of gating core on flux excursion of resetting core. The constant K depends on the resistances of the control and gating circuits

A—$\alpha \geq \gamma$
B—$\alpha < \gamma$

Vol
Pt.
195

K_3, equation 22 becomes the equation:

$$E_L=\frac{R_xR_L}{R_c(R_L+R_w)\left[\frac{NR_w}{R_L+R_w}+\frac{2R_x(I_{cm}-I_{cm}')}{\pi E_s}\right]}\times$$

$$E_c+\frac{R_xR_L\left(\frac{\pi+2}{2\pi}I_{cm}+\frac{\pi-2}{2\pi}I_{cm}'\right)}{(R_L+R_w)\left[\frac{NR_w}{R_L+R_w}+2R_x\frac{(I_{cm}-I_{cm}')}{\pi E_s}\right]}$$

or

$$E_L=K_vE_c+E_{L(o)} \qquad (22A)$$

The expression for voltage gain given in equation 22(A) provides an interesting comparison with two other expressions which may be derived from earlier developments.

If the basic difference equations 18(B) and 19(A) are reduced to the steady-state form and differentiated with respect to E_L and E_c respectively, the solution for the gain yields

$$K_v=\frac{dE_L}{dE_c}=\frac{R_xR_L}{R_c(R_L+R_w)\left[\frac{NR_w}{R_L+R_w}+\frac{2R_x(I_{cm}-I_{cm}')}{\pi E_s\sin\alpha-4\frac{I_{cm}}{N}(R_L+R_w)}\right]}$$

This result agrees with equation 22(A) at $\alpha=\pi/2$ if the I_{cm}/N term is neglected. The necessary modifications correspond exactly to those employed in the derivation of equation 21(B).

If the solution for the maximum and minimum points obtained from equations 18(B) and 19(A) is used to calculate the slope of a straight line connecting them, the result is

$$K_v=\frac{R_xR_L}{R_c(R_L+R_w)\left[\frac{NR_w}{R_L+R_w}+\frac{R_x(I_{cm}-I_{cm}')}{E_s}\right]}$$

This result is exactly that obtained if equation 21(C) were used instead of equation 21(B) to derive equation 22(A) [$(\pi/2)E_s$ replaced by E_s]. Again, the technique employs exactly the same approximations as are made in the derivation of equation 21(C). These two brief comparisons provide satisfying check points.

The approximate linear transfer characteristic as described by equation 22(A) is shown in Fig. 7. The value of $E_{c(o)}$, the intersection of the extrapolated linear transfer characteristics with the E_c axis, is

$$E_{c(o)}=\frac{E_{L(o)}}{K_v}=R_c\left(\frac{\pi+2}{2\pi}I_{cm}+\frac{\pi-2}{2\pi}I_{cm}'\right)$$

The value of $E_{c(min)}$ is determined from equation 22(A) by letting E_L be equal to the relation

$$\frac{2I_{cm}}{N}R_L$$

When this is done,[7] $E_{c(min)}$ is found to be

$$E_{c(min)}=-R_c\left\{\frac{\pi+2}{2\pi}I_{cm}+\frac{\pi-2}{2\pi}I_{cm}'-2I_{cm}\left[\frac{R_w}{R_x}+\frac{2(I_{cm}-I_{cm}')(R_L+R_w)}{\pi NE_s}\right]\right\}$$

The value of $E_{c(max)}$ is determined in a similar way by letting

$$E_L=E_s\frac{R_L}{R_L+R_w}$$

in equation 22(A). The result is the following

$$E_{c(max)}=-R_c\left[\frac{\pi-2}{2\pi}I_{cm}+\frac{\pi+2}{2\pi}I_{cm}'-E_s\frac{NR_w}{R_x(R_L+R_w)}\right]$$

The foregoing equations provide the servo designer with a quick method of predicting the approximate steady-state transfer characteristic from the circuit parameters and the core and rectifier functions. These results are indicative of small signal operation near the maximum gain region of the characteristic. For operation corresponding to large signal inputs, similar equations may be derived from equation 21(C).

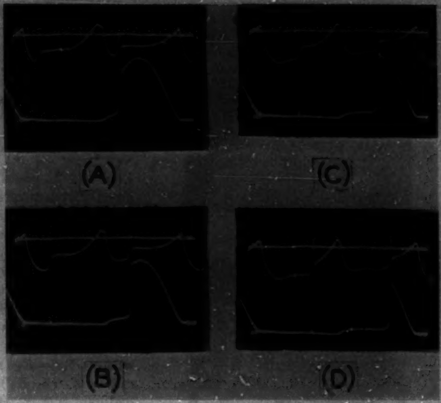


Fig. 10. Waveshapes of control and gate currents for various outputs corresponding to Fig. 8. Gate current shown at bottom of each oscillogram has a scale factor of ten times that of the control current at top

A—87½% output
B—75% output
C—50% output
D—25% output

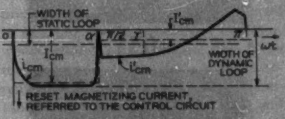


Fig. 11. Core constraints during reset of typical operating condition

DYNAMIC TRANSFER FUNCTION

The terms of equations 21(B) or 21(C) represent half-cycle average values of an assortment of wave shapes in the time domain. In order for these output difference equations to be expressed in operational form, the terms must be interpreted as Laplace transformable functions in a unique manner. One possible interpretation consists of representing the variables in the form of jump functions. With this representation, the Laplace transform of equation 21(B) becomes[9]

$$\epsilon^{Ts}[E_L(s)-e_L(O)P(s)]=K_1E_c(s)+K_2E_L(s)+K_3/s \qquad (23)$$

in which $P(s)=(1-\epsilon^{-Ts})/s$, the Laplace transform of the unit pulse, $T=\pi/\omega$, and $e_{L(o)}$ may be obtained from equation 21(B) as $e_L(O)=K_3/(1-K_2)$. Solving equation 23 for $E_L(s)$ yields

$$E_L(s)=\frac{K_1}{\epsilon^{Ts}-K_2}E_c(s)+\frac{K_3}{1-K_2}\frac{1}{s} \qquad (24)$$

Equation 24 contains all of the pertinent information concerning the dynamic operation of the magnetic amplifier, provided the limitation imposed by the interpretation of E_L and E_c as jump functions is recognized. Although this limitation is discussed in more detail elsewhere,[10]

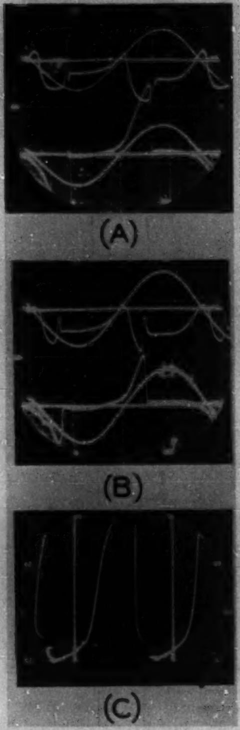


Fig. 12. Effect of control resistance on I_c, e_{core}, and flux reset

A—R_c small
B—Flux reset
C—R_c large

if this representation is used in the system with other frequency-sensitive networks, then it is inherently assumed that the jump function is a good approximation of the input and output time functions. In other words, the input frequencies are small compared with the supply-voltage frequency, and the load response to a jump function is essentially the same as to the actual output. These conditions are invariably fulfilled in practice. Of course, if the magnetic amplifier is used alone or if the appropriate half-cycle integrators and other appropriate circuitry are used, the representation becomes exact in theory.

With the previous procedure and discussion applicable, the terms of equation 21(B) may be interpreted as impulse functions and z-transformed.[11]

$$z[E_L^*(z)-e_L(O)]=K_1E_c^*(z)+K_2E_L^*(z)+K_3z/(z-1) \quad (25)$$

with the value of $e_L(O)$ as before,

$$E_L^*(z)=\frac{K_1}{z-K_2}E_c^*(z)+\frac{K_3}{1-K_2}\,\frac{z}{z-1} \quad (26)$$

Equation 26 could have been obtained directly from equation 24, which contains the same information.

Examination of equation 26 yields the following information: the response to a step input is exponential with a time constant of $\tau=-T/\ln K_2$; the steady-state gain is $K_1/(1-K_2)$; in the absence of signal the output is $K_3/(1-K_2)$; the frequency-response locus in the $G^*(z)$ plane, where $G^*(z)=K_1/(z-K_2)$, is a circle of radius $K_1/(1-K_2^2)$ and center on the real axis at $K_1K_2/(1-K_2^2)$.

If the first two terms of the series expansion of $\ln K_2$ for $K_2 \cong 1$ are substituted into the expression for time constant

$$\tau\cong\frac{\pi}{\omega}\,\frac{R_x+N^2R_c}{N^2R_c\left[\dfrac{R_w}{R_L+R_w}+\dfrac{2R_x\left(\dfrac{I_{cm}}{N}-\dfrac{I_{cm}'}{N}\right)}{\pi E_s}\right]} \quad (27)$$

This procedure is equivalent to the substitution of $e^{Ts}=1+Ts$ into equation 24. Equation 27 together with one of the steady-state gain expressions gives quick although approximate insight into the magnetic-amplifier dynamic characteristics. The approximate gain-bandwidth product K_v/τ remains K_1/T as in the case with assumed negligible width of the B-H loop.

Square-Wave Excitation

Assumption of an alternating square-wave excitation in the gate circuit and the corresponding direct-voltage excitation in the control circuit may be deduced quickly and easliy from the more complicated situation examined. Modes A and B no longer exist since sufficient excitation voltage is immediately available at the beginning of each half-cycle. The voltage terms e_s and e_c in the equations describing modes C and D become the constant values of the gate and control voltages during a half-cycle of the alternating voltage, and therefore are identical with the half-cycle average voltage symbols E_s and E_c.

The determining equation in terms of voltages, equation 18(B), remains the same except that the terms providing the proper proportionality factor for a fraction of the half-cycle, $(1-\cos\alpha)/2$ and $(1+\cos\alpha)/2$, become α/π and $1-(\alpha/\pi)$ respectively. The same modification applies to the voltage, equations 19(A) and 20(A). These three equations are linear in the parameters $\alpha(n)$ and $\alpha(n+1)$ instead of transcendental and may be solved without any approximations. If the contribution of the magnetizing currents to the load current is neglected ($I_{cm}R_L$, $I_{cm}R_w << E_s$), the solution for $E_L(n+1)$ is exactly that of equation 21(C), certainly an expected result because of the approximation used to obtain the equation originally. All conclusions deduced from equation 21(C) apply to the case of square-wave excitation with direct-voltage control.

Dynamic Behavior of the B-H Loop

The exact behavior of the cores at all points during a cycle of operation is very complicated and not amenable to reasonable analytic treatment. An attempt to formulate an accurate analytic representation has appeared.[8] However, any exact formulation is inherently difficult to incorporate into the already complex description of the behavior of the magnetic amplifier. Empirical methods will probably not be satisfactory because of the apparent effect of a wide variety of factors with the attendent loss of generality, and understanding will probably come only as deeper insight is gained into the ferromagnetic phenomena themselves. Fortunately, however, there are reasonable analytic approximations to the core behavior which give practical results.

FLUX CHANGE DURING MODE D

The region of chief interest in the formulation of a real-core function occurs during the output period when one core is saturated and the other unsaturated, mode D from α to π. During this mode the flux level of the resetting core continues to change although at a smaller rate than during mode C. This smaller rate leads to the common postulation that operation is along the static loop instead of the dynamic loop as in mode C. Fig. 8 shows the behavior of an unsaturated core under actual operating conditions for several values of output. The noticeable oscillations at firing are not of interest here and are discussed in detail in another report.[8] It is interesting to note that the magnitude of flux change during mode D for each of the outputs is nearly the same. This fact is explained by equations 14 and 16 and by realizing that the flux change is dependent on the size of the integrand and on the limits of the integral, both of which are functions of e_c. In this case

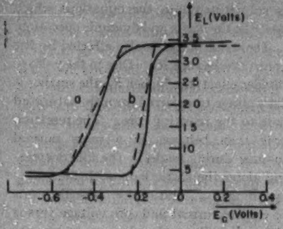


Fig 13. Static transfer characteristic with $R_w=350$ ohms, $R_L=5{,}000$ ohms, $R_b>10^8$ ohms. Solid curves are experimental, dashed curves calculated

a—$R_c=150$ ohms b—$R_c=64$ ohms

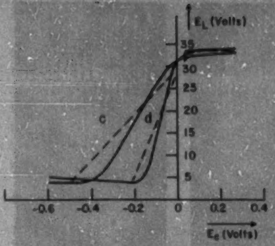


Fig. 14. Static transfer characteristic with $R_w=350$ ohms, $R_L=5{,}000$ ohms, $R_b=10^4$ ohms. Solid curves are experimental, dashed curves calculated

c—$R_c=150$ ohms d—$R_c=64$ ohms

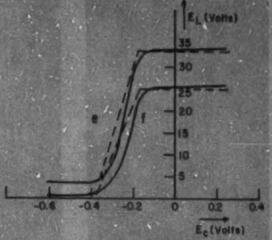


Fig. 15. Static transfer characteristic with $R_c=100$ ohms, $R_b>10^8$ ohms. Solid curves are experimental, dashed curves calculated

e—$R_L=5{,}000$ ohms, $R_w=350$ ohms
f—$R_L=500$ ohms, $R_w=220$ ohms

the two effects oppose. However, the relative contribution of the flux reset during mode D to the total flux reset becomes more important as the output increases.

At the two extreme points of maximum and minimum output, $\Delta\phi$ during mode D was predicted to be zero in the analysis. Fig. 8 shows that this is correct in the case of minimum output when the limits of integration coincide. However, when α approaches β_3 the assumption of a square *B-H* loop is remote from reality. Furthermore, the saturation region obviously has a nonzero slope which introduces a small inductance L_{sat}. As a result, it is quite hard to locate the exact point of the maximum output. The same conclusion can be reached from the transfer characteristics in which no clear-cut maximum output point is located.

In general, as $\alpha\rightarrow\pi$, the closer is the assumption of a square *B-H* loop to reality. The portions of heavier line in Fig. 8 correspond to a small change of flux during a comparatively longer period of time. Note that these heavy lines extend beyond the minimum point of the flux. It is of interest to discuss this phenomenon, and to investigate the flux changes around the minimum flux point of the unsaturated core.

EFFECT OF SATURATION INDUCTANCE

The assumption that the saturation region is absolutely horizontal is not fulfilled in practical cores. Fig. 8 shows that there is some small but finite slope in this region. This continued coupling between the gate and control windings of the gating core has a definite effect on the resetting core.

When the gating core fires at $\omega t=\alpha$, the flux of the gating core continues to increase; the voltage associated with this flux change remains positive until the core reaches the point of maximum excursion into the saturated region. A fraction of this voltage appears as an additional component across the resetting core, tending to decrease the flux as seen in Fig. 1. As the gating core returns from its excursion into the saturated region the effect changes polarity, producing a voltage across the resetting core which tends to increase the flux. These voltages transferred from the gating to the resetting core, and with the normal mode D voltage, which tends to decrease the flux, determine the complicated behavior of the resetting core.

At some point $\omega t=\gamma$, the effect of the transferred voltage and the normal mode D voltage across the resetting core cancel, and at this point the maximum reset flux is reached. From the polarity of the voltages involved, $\gamma>\pi/2$. As the firing angle α becomes larger than $\pi/2$ it may also become greater than γ, i.e., the transferred voltage may immediately cancel the normal mode D voltage and the flux of the resetting core may begin to rise at α with no further reset from α to π. These phenomena may be readily observed in Fig. 8. For example, at 75% full output $\alpha<\gamma$, whereas at 25% full output $\alpha>\gamma$.

Fig. 9 summarizes the effects of the saturated inductance, which is still assumed small enough so as to produce negligible lag between the load current and load voltage. The important fact to be noted is that the net flux reset between α and π remains $\Delta\phi_{rhc,D}$, as developed in the analysis. The saturated inductance of the gating core makes no net contribution to the total flux change of the resetting core. However, because of the nonlinear properties of the core, the mmf requirements, i.e., i_{cm}', are definitely and appreciably affected.

CORE FUNCTIONS I_{cm} AND I_{cm}'

The analysis was based on the assumption of a rectangular *B-H* loop, and expressions to determine α were developed in terms of I_{cm}, I_{cm}', and E_c. However, when dealing with practical cores, the magnitudes of I_{cm} and I_{cm}' should represent the average values of the instantaneous constraints of the core, i_{cm} and i_{cm}'. Hence

$$I_{cm}=(1/\alpha)\int_0^\alpha i_{cm}d\omega t$$

and

$$I_{cm}'=(1/\pi-\alpha)\int_\alpha^\pi i_{cm}'d\omega t$$

With reference to Fig. 8, it is obvious that the actual constraints of the core result in rounded corners on the *B-H* loop in both places where the core starts to proceed towards saturation and where it reaches saturation. These two rounded corners tend to cancel each other when the average value of i_{cm} in the front side of the loop is calculated. On the left side of the loop the rectangular *B-H* loop is rounded mostly where the core starts resetting from saturation. The effect is small at minimum output and I_{cm} may be taken as the dynamic coercive force, whereas at smaller firing angles I_{cm} is slightly smaller. However, except for the region close to the maximum output, the coercive force of the dynamic loop may still be regarded to a good engineering approximation as I_{cm}.

The waveshape of i_{cm}' is entirely different because of the effect of saturation inductance. The oscillograms of Fig. 8 are not sufficient for investigation because the time factor is missing there. Oscillograms of i_{cm} and i_{cm}' corresponding to Fig. 8 are shown in Fig. 10. Typical waveshapes of i_{cm} and i_{cm}' are presented graphically in Fig. 11. From α to γ (provided

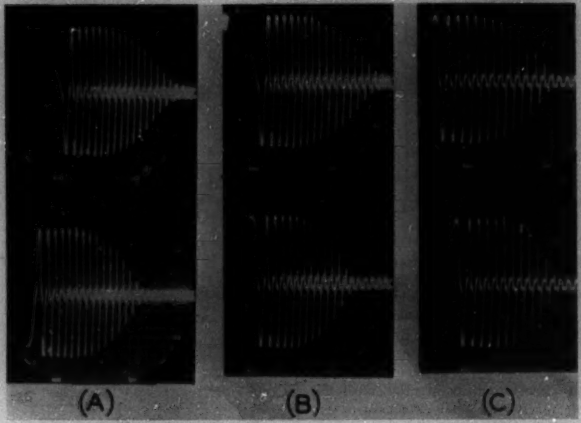


Fig. 16. Transient response to a step of control voltage resulting in change of output from $E_L(E_c=0)$ to $E_{L\min}$. Results are given in half-cycles m of the supply frequency with calculated values in parentheses

A—$R_b>10^6$ ohms; top, $R_c=100$ ohms, m=16(19); bottom, $R_c=400$ ohms, m=10(5)
B—$R_b=50{,}000$ ohms; top, $R_c=100$ ohms, m=16(18); bottom, $R_c=400$ ohms, m=10(7)
C—$R_b=5{,}000$ ohms; top, $R_c=100$ ohms, m=12(16); bottom, $R_c=400$ ohms, m=12(12)

$\alpha<\gamma$) the flux continues to reset, and i_{cm}' is approximately constant. At $\omega t>\gamma$ the resetting voltage changes polarity due to the saturation inductance, the flux begins to increase, and i_{cm}' is reduced. In this case investigation of Figs. 8 and 10 reveals that γ is slightly larger than $\pi/2$. The current i_{cm}' reaches the value of zero somewhere between γ and π, and then changes sign. It is safe to conclude that I_{cm}' is a function of α.

If the waveshapes of Fig. 10 are compared with the magnetizing curves given by the core manufacturer, it is concluded that I_{cm}' decreases as $\alpha\rightarrow\pi$, and may be smaller than the width of the static loop, as shown in Fig. 11. These conclusions are very important when the value of I_{cm}' is substituted into the equations which describe the static and dynamic operation.

The value of I_{cm}' is also affected by the control resistance R_c. Due to the voltage divider effect, the higher R_c, the smaller is the voltage transferred from the saturated core to the resetting core. The resetting core then behaves in a more normal manner during mode D, the flux decreasing along the static loop with the corresponding i_{cm}' and $\gamma\rightarrow\pi$. Oscillograms of control current and core voltage versus time as well as the flux excursion are shown in Fig. 12 for two values of R_c. In both cases the same average control current I_c exists. When R_c is large, α becomes smaller since the reset from $\omega t=0$ to $\omega t=\alpha$ is smaller. On the other hand, the effect of the saturation inductance on decreasing the magnitude of I_{cm}' is reduced.

I_{cm}' is also affected by the load resistance. When R_L becomes smaller the saturation current is higher. As a result, a higher voltage $(L_{sat})di/dt$ is induced into the control circuit, causing i_{cm}' to be more positive, and the magnitude of I_{cm}' becomes smaller.

This discussion serves to illustrate the complexity which any "exact" analytic representation of the behavior of the resetting core would assume. If test results from actual operation are available, then I_{cm}' may be chosen reasonably accurately for the dynamic transfer function. The designer must choose I_{cm}' as the width of the static loop for want of a better approximation, although an empirical decrease of 10% will give good results with many circuits and core materials.

Experimental Results

The experimental results are obtained with a magnetic amplifier which has the following design characteristics:

Deltamax cores *4179D1*
$N_G=5{,}000$ turns no. 33
$N_c=500$ turns no. 33
$H_cl_c=4.4$ milliamperes (dynamic referred to control circuit)
$H_cl_c=1.2$ milliamperes (static referred to control circuit)
$e_s=36\ (\pi/2)\sin 2\pi\, 60t$ volts
$R_L=5{,}000$ ohms

The control resistance R_c and the rectifier back resistance R_b are varied to produce different dynamic characteristics. Equation 21(A) and even more accurate expressions which do not involve the approximations $R_s>>R_w$, R_L, and $e_s>>i_{cm}R_w$, $i_{cm}R_L$, will give remarkable correlation between experimental and calculated steady-state characteristics for a wide range of R_c and R_b.[8] However, the pur-

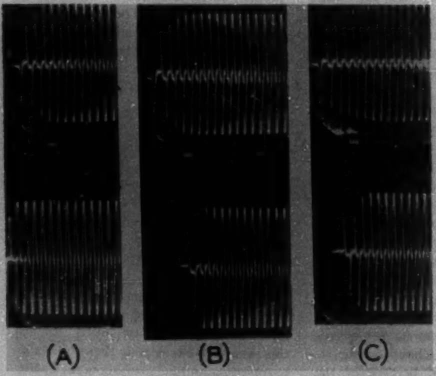


Fig. 17. Transient response to a step of control resulting in change of output from 25% to E_L ($E_c=0$). Results are given in half-cycles m of the supply frequency with calculated values in parentheses

A—$R_b>10^6$ ohms; top, $R_c=100$ ohms, m=18(19); bottom, $R_c=400$ ohms, m=5(5)

B—$R_b=50{,}000$ ohms; top, $R_c=100$ ohms, m=16(18); bottom, $R_c=400$ ohms, m=7(7)

C—$R_b=5{,}000$ ohms; top $R_c=100$ ohms, m=16(16); bottom, $R_c=400$ ohms, m=12(12) pose of the experimental results here is to verify the results obtained from equation 21(B) from which the gain and time constant are obtained. Figs. 13 through 15 compare calculated and experimental results and show reasonable agreement between steady-state gain characteristics. In large signal applications the magnetic-amplifier transfer function must be supplemented by limiting in order to represent the full and minimum output conditions.

The oscillograms of Figs. 16 and 17 verify the validity of the theoretical expression for time constant.

Conclusions

The dynamic operation of the magnetic amplifier may be described by an output difference equation which contains realistic approximations of real-core and rectifier functions. The nonlinear difference equation may be linearized in order to obtain a useful transfer function in operational form. The results predicted by the operational transfer function may be verified experimentally for a range of values of control-circuit resistance and rectifier back resistance.

References

1. Analysis of Magnetic Amplifiers by the Use of Difference Equations, P. R. Johannessen. *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 700–11.

2. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955.

3. The Operation of Magnetic Amplifiers with Various Types of Load—Parts I, II, L. A. Finzi, R. R. Jackson. *AIEE Transactions*, vol. 73, pt. I, July 1954, pp. 270–88.

4. On the Mechanics of Magnetic Amplifier Operation, Robert A. Ramey. *Ibid.*, vol. 70, pt. II, 1951, pp. 1214–23.

5. Design Considerations of the Half-Wave Bridge Magnetic Amplifier, C. W. Lufcy, H. H. Woodson. *Ibid.*, vol. 73, pt. I, July 1954, pp. 229–26.

6. Preliminary Analysis of the Half-Wave Bridge Magnetic Amplifier, H. H. Woodson. *Report 3596*, U.S. Naval Ordnance Laboratory, Silver Springs, Md., Dec. 1953.

7. Dynamic Operation of Magnetic Amplifiers for Feedback Control Systems, H. C. Bourne, D. Nitzan. *TN-56-555*, Air Force Cambridge Research Center, Air Research and Development Command, Cambridge, Mass., May 1956.

8. Dynamic Operation of Magnetic Amplifiers for Feedback Control Systems, H. C. Bourne, T. T. Kadota, D. Nitzan. *TN-56-769*, *Ibid.*

9. Transients in Linear Systems (book), M. F. Gardner, J. L. Barnes. John Wiley & Sons, Inc., New York, N. Y., 1942, p. 289.

10. Operational Magnetic Amplifiers with Multiple Control Windings, T. T. Kadota, H. C. Bourne. *AIEE Transactions*, vol. 76, pt. I, 1957 (*Paper no. 57-772*).

11. Analyses and Synthesis of Sampled-Data Control Systems, Eliahu I. Jury. *Ibid.*, vol. 73, pt. I, Sept. 1954, pp. 332–46.

Microwave Radio Communication Under Severe Winter Conditions on the Bonneville Power Administration System

E. M. STRONG
NONMEMBER AIEE

E. J. WARCHOL
MEMBER AIEE

S. E. BENSON
NONMEMBER AIEE

THE MICROWAVE STATIONS operating in the mountainous regions of the Pacific Northwest are subjected to severe winter conditions. Conventional antenna and tower design practices could not be used in the Bonneville Power Administration System.

Three natural problems were recognized for successful transmission of the microwave beam from point-to-point in the mountain ranges operating under severe winter conditions:

1. The microwave antenna towers must be of rugged design to withstand the ice, snow, and wind loading.
2. The microwave antenna reflectors and antenna feed horns or dipoles must be protected from the accumulation of ice and snow, and protected against damage caused by falling ice.
3. Passive repeaters (plane reflectors) are used advantageously to reduce the cost of the system and permit repeater stations to be located at accessible sites.

The Cascade Mountain Range is traversed by multichannel microwave in two places: first, through Snoqualmie Pass in central Washington, and second, through the Columbia River Gorge. A third location where special design was required for successful winter operation of a microwave station is on top of Mt. Spokane. Mt. Spokane, located in eastern Washington and 23 miles northeast of the city of Spokane, is ideally situated in relation to the system. A road and a 3-phase distribution power line existed to the top of the site. Although the elevation of Mt. Spokane, 5,800 feet, is not relatively high, the facilities at the site, which include a microwave radio station, a television broadcasting station, a mobile VHF (very high frequency) station, a state-forest lookout tower, and ski facilities, are subject to unusual winter storms.

General System Layout

The headquarters for dispatching the electrical power generated at the federal hydroelectric plants is located in the United States Department of the Interior at Portland, Oreg. Since many different types of communication circuits are needed for the dispatching facilities, Portland is also the central terminal station for the communication system. Microwave radio is used to control the megawatts of generated power. The multichannels provide circuits at Portland for dispatching, telemetering, water stage telemark, supervisory control, VHF mobile communications, private-automatic-exchange tie trunks, power-line-carrier tie trunks, and teleprinting of remote quantities. Fig. 1 shows the interior building, and Fig. 2 is a close-up view of the microwave antenna tower. The two 6-foot-diameter and two 10-foot-diameter parabolic antenna reflectors direct the microwave radio beam to the north, south, east, and across the Columbia River to the J. D. Ross Substation which is near Vancouver, Wash. The 23 channels in the Portland to Ross Microwave System are used to extend power-line-carrier telephone-dispatching circuits, for telemetering frequency, voltage, and load quantities, and to provide tie trunks between the automatic telephone switchboards. General data for the microwave radio-communication system shown on Fig. 3 are given in Table I.

Paper 57-915, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Pasco, Wash., August 28–30, 1957. Manuscript submitted May 31, 1957; made available for printing July 5, 1957.

E. M. Strong and E. J. Warchol are with the Bonneville Power Administration, Portland, Oreg.; S. E. Benson is with the Philco Corporation, Philadelphia, Pa.

Fig. 1. Portland interior building

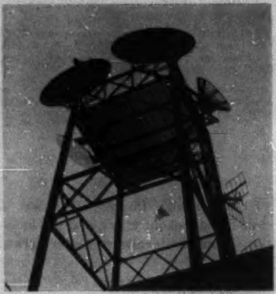


Fig. 2. Close-up view of microwave antenna tower on top of the interior building

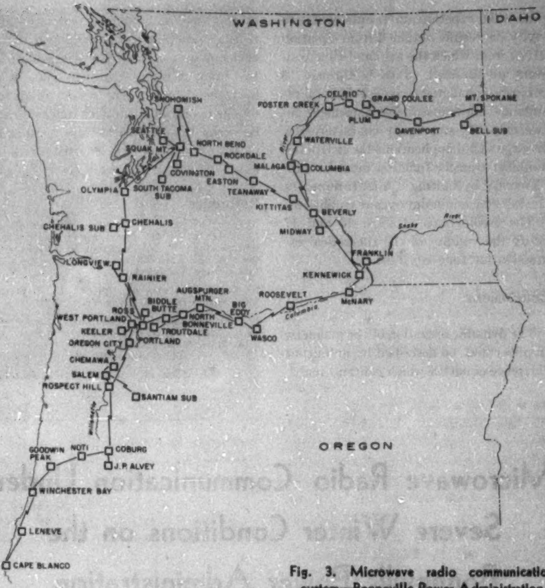

WASHINGTON
IDAHO
OREGON
SNOHOMISH
SEATTLE
SQUAK MT.
NORTH BEND
ROCKDALE
COVINGTON
EASTON
TEANAWAY
OLYMPIA
SOUTH TACOMA SUB
CHEHALIS SUB
CHEHALIS
LONGVIEW
RAINIER
FOSTER CREEK
DELRIO
GRAND COULEE
PLUM
DAVENPORT
MT. SPOKANE
BELL SUB
WATERVILLE
MALAGA
COLUMBIA
KITTITAS
BEVERLY
MIDWAY
River
Snake
River
FRANKLIN
KENNEWICK
McNARY
ROOSEVELT
Columbia
AUGSPURGER MTN.
BIDDLE BUTTE
BIG EDDY
ROSS
WEST PORTLAND
NORTH BONNEVILLE
TROUTDALE
KEELER
OREGON CITY
PORTLAND
WASCO
CHEMAWA
SALEM
PROSPECT HILL
SANTIAM SUB
Willamette
GOODWIN PEAK
NOTI
COBURG
J.P. ALVEY
WINCHESTER BAY
LENEVE
CAPE BLANCO

Fig. 3. Microwave radio communication system—Bonneville Power Administration

Description of Terrain and Winter Conditions of Selected Station Sites and Routes

North and South Microwave System

The stations in the North and South system are situated in the valleys between the Cascade and Coastal Mountain Ranges. They are not subjected to unusually heavy ice and snow conditions. The elevation of the highest station, Squak Mountain, is 1,980 feet. It is situated in the Puget Sound area to the north, bordering on the western slope of the Cascade Mountain Range. The 6-foot and 10-foot diameter parabolic antenna reflectors and dipole antennas that are used for frequencies between 1,700 and 1,850 megacycles (mc) are unprotected. Heaters or protective covers are not used for two reasons: (1) satisfactory operation is obtained with up to 2 inches of ice or snow covering the reflector and dipole antenna, and (2) the stations are not subjected to severe winter climatic conditions. Fig. 4 shows the Squak Mountain Repeater Station. Because of the large number of antennas used, two self-supporting steel towers each 54 feet high were required. One antenna is connected to both transmitter and receiver operating in the same direction by means of duplexing filters. The duplexing feature reduces the number of antennas and tower space required. Radio-frequency stability is important in the Squak Mt. area because of the frequency congestion. The large number of power-system control circuits and telephone traffic are handled by eight primary transmitters with stand-by and eight receivers operating in five different directions.

East Microwave System

Large blocks of power generated on the Columbia River are brought to the market through the Columbia River Gorge. Bonneville Dam is situated in the deepest part of the Columbia River Gorge where it cuts through the Cascade Mountain Range. Two main highways, two railroads, and several high-voltage transmission lines follow this route to the coast. From Portland, the East Microwave System follows the 230- and 345-kv transmission lines up the river and connects with the substations and hydroelectric generating stations. This is the first microwave installation to cross the Cascade Mountain Range in the Pacific Northwest. Two repeater stations and one way station are used to relay the signals through the Columbia River Gorge. These stations are Biddle Butte Repeater Station, 22 miles east of Portland, Bonneville Dam Way Station, and Augspurger Mountain Repeater Station.

Table I. Microwave Radio Communication System Data, May 28, 1957

System	No. of Stations	Total Length, Miles	No. of Channels	Equipment Manufacturer	RF Frequency Band, Mc
North	9	245	23	FTR*	1,700–1,850
	4	12	5	Motorola	7,125–7,425
South	10	170	23	FTR*	1,700–1,850
	6	130	1	Philco	7,125–7,425
East	30	728	24	Philco	7,125–7,425
J. D. Ross	2	9	23	FTR*	1,700–1,850

* FTR—Federal Telephone and Radio Corporation.

Fig. 4. Squak Mountain microwave repeater station

Fig. 5. Augspurger Mountain microwave parabolic antenna reflector, antenna feed, and tower covered with frost, Feb. 5, 1954

Biddle Butte Repeater Station connects Troutdale Substation to the main microwave system by a branching spur. Although this is situated at a relatively low elevation, 1,900 feet, one or two sleet storms during the winter will encrust the antennas and tower with 2 inches or more of solid ice. Similar conditions are encountered at the next station, Bonneville Dam, which is situated on Bradford Island adjacent to the 350-foot river-crossing towers. In addition to sleet protection the antennas mounted on the roof of Bonneville Dam microwave building must be protected from solid chunks of ice falling from the 230-kv transmission line conductors and ground wires on the river crossing towers. The river crossing towers are used to support two plane reflectors, one 250 feet up the tower facing west towards Augspurger Mountain Station. The two roof-mounted parabolic antenna reflectors are protected from falling ice by cyclone wire fencing secured to a steel frame. The falling ice shatters into small pieces or glances off to one side. The attenuation to the r-f (radio-frequency) signal is less than 1/2 db (decibel) which is small.

Augspurger Mountain is above the Columbia River at an elevation of 3,200 feet on the east side of the Cascade Mountain Range. During a winter storm the prevailing east winds will drive snow against the microwave antennas. The clinging snow resembling frost will build out to 12 inches or more and block the r-f energy. Fig. 5 shows typical conditions and what can happen to unprotected antennas. The 1-kw Calrod heating element covering an area of 2 feet in diameter on the back of the 6-foot-diameter parabolas failed to keep the face of the parabola clear. A 20-watt heater on the tip of the wave-guide feed horn also froze over.

The parabolic antennas and their feed systems are particularly sensitive to the weather. The performance of the equipment depends to a great extent upon maintaining the antennas at their peak of performance. Anything which will distort the shape of the parabola or move the position of the feed, with respect to the focus of the parabola, will reduce its gain or alter its beam angle. Furthermore, foreign material, ranging from ice and snow to insects and birds' nests, will alter the field pattern and impair the performance.

While the ice shields over the antennas previously described are of tremendous advantage from the standpoint of affording protection against the impact of falling ice, additional mechanical protection is necessary. The first attempts at this

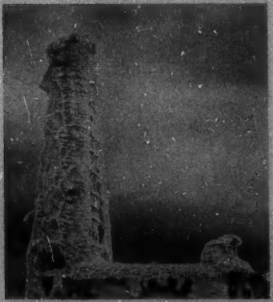
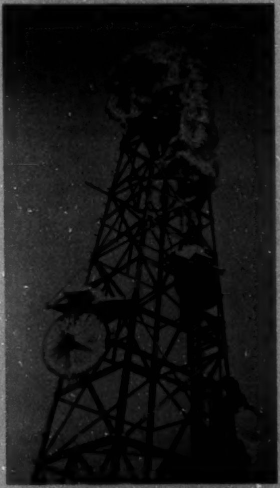


Fig. 6. Mt. Spokane microwave tower and antennas covered with snow



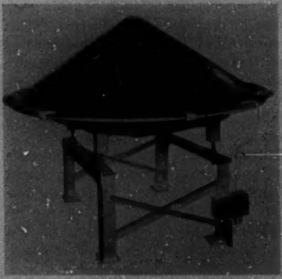


Fig. 7 (left). Mt. Spokane microwave tower and antenna reflectors with fabric covers and ice shields

Fig. 8 (above). Close-up of thermo-microdome

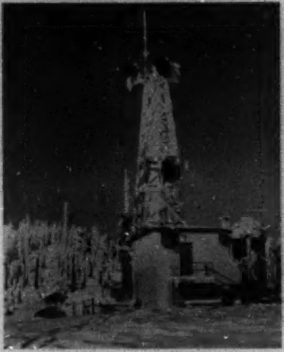


Fig. 9. Augspurger Mountain microwave station with thermo-microdome protective covers

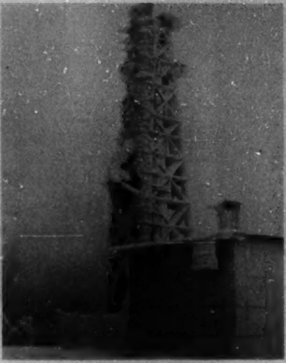


Fig. 10. Mt. Spokane microwave tower and antenna reflector protected with thermo-microdome and ice shield. Circuit is working

were to apply thermostatically controlled Calrod heating elements to the back of the parabola to melt off any snow or ice which might form. The heat transmission to the front surface with this arrangement was relatively poor since much of the heat was lost through back radiation and dissipation through the supporting structure. It was found that it was almost impossible to provide heaters of sufficient capacity to protect the antennas under the extremely heavy icing conditions which prevail in the Pacific Northwest area. Even though heating elements of enormous capacity were provided, the ice bridged over the front of the antenna, as shown in Fig. 6, and destroyed its effectiveness. This method, of course, did nothing to protect against the deposit of other materials than snow and ice.

A cone-shaped cover of vinyl-impregnated nylon fabric was then tried. This was fitted over the parabola and clamped at the edges. Covers of this type are shown in the view of Fig. 7. The theory was that the natural movement of the fabric in the wind would be sufficient to break off any ice which might form. This appeared to work with small amounts of icing but again it was overpowered under heavy icing conditions. The "tooth" of the fabric was sufficient to allow some ice to form. This stiffened the fabric and the icing could then continue to grow unimpeded.

The successful solution to the problem employed the brute force approach of covering the entire front surface of the antenna with a suitably stiff protective enclosure with special design requirements.

It was realized that such a cover would require high impact strength, negligible attenuation to the signal, a smooth surface providing a minimum of tooth to which the ice and snow might adhere, an integral heating element to melt ice and snow, and simple attachment means. A structure was designed of a conical shape, as shown in Fig. 8, of a size to fit snugly against the edge of the parabola. The material chosen was a 3-ply fiber-glass fabric laminated with polyester resinate. This material has high impact resistance, as is evidenced by its popular use for heavy-duty containers and even for the hulls of boats. To take advantage of the maximum absorption effect of the sun's heat, a dark color, olive drab, was selected. While black would have been better, it was feared that black pigmentation, generally of a carbonaceous material, would adversely affect the transmission of the microwave signal. Resistance wire to provide suitable heating capacity was imbedded in the plastic material. The pattern of the wires was carefully arranged to minimize the signal attenuation. Exhaustive laboratory tests were made under simulated icing conditions to determine the optimum heating capacity which would be required. Finally, a clamping arrangement was designed which would permit easy attachment of the cover to existing antenna structures even while mounted on towers.

The final design was applied to three sizes of parabolas: 4, 6, and 8 feet in diameter, with heating elements rated at 1,250, 2,500, and 4,550 watts respectively. High- and low-temperature thermostats were provided to apply power to the heating elements while the ambient temperature would be in the range of approximately +10 to +40 degrees Fahrenheit (F). Heat is applied above freezing temperatures to prevent the possibility of any icing beginning to form since it takes more heat to melt ice once it has formed than is required to keep it from forming. At temperatures below +10 F there is seldom, if ever, the type of precipitation which forms heavy icing and so the power is turned off. Measurements of the final design installed on antennas show no effect on the voltage standing wave radio and less than 0.1 db attenuation so that the cover can be considered completely transparent to the microwave signal.

Thermo-microdomes, as these covers

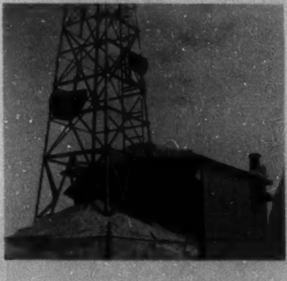


Fig. 11. Mt. Spokane tower and antenna reflectors during construction with no protection from falling ice

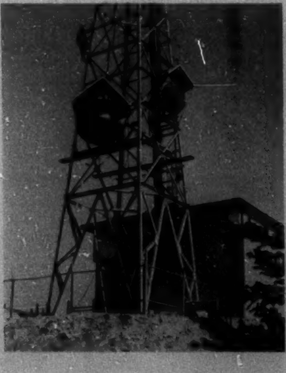
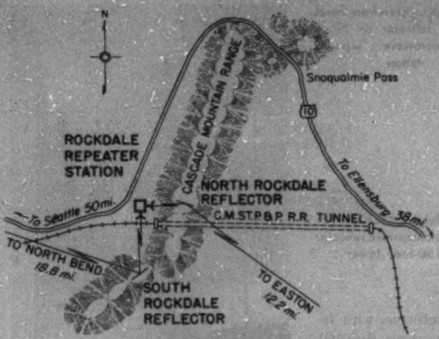


Fig. 12. Mt. Spokane microwave tower and antenna reflectors with thermo-microdomes and new type of ice shields with vibrators


N
Snoqualmie Pass
10
CASCADE MOUNTAIN RANGE
ROCKDALE
REPEATER
STATION
NORTH ROCKDALE
REFLECTOR
To Seattle 50mi.
C.M.ST.P.&P. R.R. TUNNEL
To Ellensburg 38mi.
TO NORTH BEND
18.8mi.
SOUTH
ROCKDALE
REFLECTOR
TO EASTON
12.2mi.

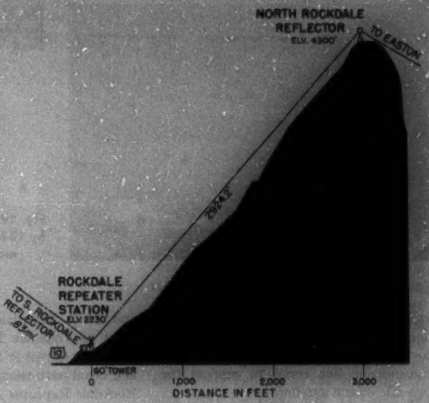

NORTH ROCKDALE
REFLECTOR
ELV. 4300'
TO EASTON
2924.2'
ROCKDALE
REPEATER
STATION
ELV. 2230'
TO S. ROCKDALE
REFLECTOR
.83mi.
10
60' TOWER
0
1,000
2,000
3,000
DISTANCE IN FEET

Fig. 13 (above). Rockdale microwave beam paths

Fig. 14 (right). Profile Rockdale repeater to North Rockdale reflector beam path

are called, were installed on all the antennas in the troublesome weather locations. The installation at Augspurger Mountain, shown in Fig. 9, is typical. This photograph dramatically indicates the effectiveness of the covers under heavy icing conditions. Fig. 10 also shows successful operation during a storm on top of Mt. Spokane. The snow cat seen near the entrance to the building in Fig. 6 is used during the winter months. The access road is also used for maintenance of the 230-kv transmission line.

It has been stated that the type of winter conditions encountered on top of Mt. Spokane is not found anywhere else in the United States. The freezing snow will build solidly around the tower as much as 8 feet. The tower supporting the microwave parabolic antenna reflectors, whip antennas for VHF mobile communications and corner reflector UHF (ultra high frequency) antennas for point-to-point communications to Albeni Falls Dam is a heavy-duty type. The damage that can be expected by falling snow and ice to parabolic antenna reflectors with no ice shields is seen in Fig. 11. This photograph was taken shortly after the first winter storm arrived before ice shields could be installed. A new type of ice shield equipped with pneumatic vibrators and thermo-microdomes is shown in Fig. 12. No circuit outages have been encountered during the winter of 1956 to 1957 with this type of protection.

Fig. 15. North Rockdale reflector and tower

Another section of the East Microwave Radio Communication System crosses the Cascade Mountain Range again at Snoqualmie Pass. Snoqualmie Pass is one of the main outlets via a railroad and a new 4-lane highway from eastern Washington to Seattle, the main seaport on the Puget Sound. The terminal stations for this section of the microwave system are located at Snohomish Substation which is north of Seattle and Columbia Substation in eastern Washington near Wenatchee.

In addition to providing essential communication circuits for the power transmission network in this area this section provides alternate circuits to the hydroelectric generating stations, substations, and the utilities connected in the Northwest Power Pool operating in eastern Washington.

The snow depth at the two repeater stations on both sides of Snoqualmie Pass, Rockdale and Easton, will exceed 20 feet. Two story buildings were selected for these two sites as for Augspurger Mountain and Mt. Spokane in order readily to enter the building through the second floor door. The main power panels and emergency power equipment, consisting of a flywheel motor-generator set and a propane engine-generator unit in a two story building, are installed on the first floor and microwave and VHF mobile-radio equipment are installed on the second floor. Inlet and outlet ports for the building and engine ventilating duct-work are mounted higher above ground to prevent snow from plugging them. Easton is situated on a prominent knoll in the

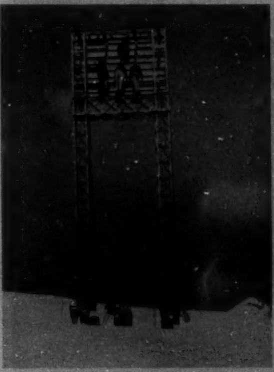


Fig. 16. South Rockdale reflector and tower

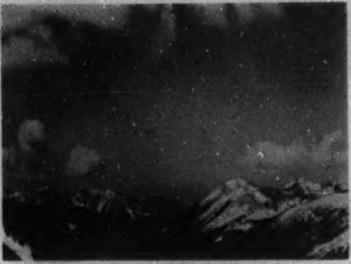


Fig. 17 (left). View from South Rockdale reflector to North Bend microwave repeater station

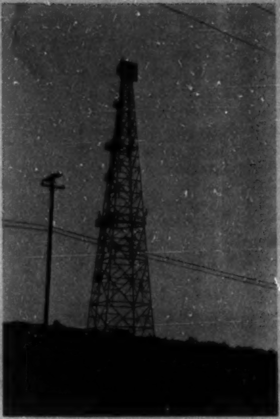


Fig. 19 (right). McNary 10- by 10-foot microwave reflector and 150-foot tower

valley adjacent to a 230-kv transmission-line tower. It was possible to save the cost of a tower 100 feet high by using the transmission line tower to support the two microwave antennas.

Application of Passive Repeaters

A unique application of two passive repeaters has been made at Rockdale to deflect the microwave beam over Snoqualmie Pass. This has made it possible to locate Rockdale Repeater Station next to United States Highway 10 near the summit of the pass and also made it possible to beam to adjacent repeater stations via plane reflectors located on top of two different peaks. No one peak existed in the area that was relatively low in elevation and accessible that commanded a line-of-site to both adjacent repeater stations. Locating the microwave station near the new 4-lane highway eliminated a costly access road.

A geographical layout of the North and South Rockdale reflectors with respect to Rockdale Microwave Repeater Station is shown in Fig. 13. The profile and beam path that was cleared between Rockdale Repeater and North Rockdale Reflector Station is shown in Fig. 14. Figs. 15 and 16 are views of North and South Rockdale Reflectors respectively. These photographs were taken in April of 1956 following a heavy accumulation of snow during the winter. The snow around South Rockdale reflector at the time this photograph was taken was heavily packed.

The microwave beam between North Bend Repeater Station and South Rockdale Reflector passes through a *V* formed by mountain ranges on both sides of the valley. A view of this 18.8 mile transmission path to North Bend is seen in Fig. 17. The path from South Rockdale Reflector to Rockdale Repeater Station is seen in Fig. 18.

A 3-reflector system is employed in the East Microwave Communication System between McNary Substation and Kennewick Repeater Station. Two reflectors are used, one at each of the two microwave stations. A reflector is mounted on a self-supporting steel tower 100 feet high at each station. The third reflector is located 0.59 of a mile from McNary Substation across United States Highway 730. The plane surface is 0.081 solid aluminum plate riveted to a 10-foot square aluminum frame. Fig. 19 is a reflector installed on top of the 150-foot tower.

Another more complicated reflector scheme was required between Wasco Microwave Repeater Station and Big Eddy Substation. This is another case where a good substation site is not a good microwave station site. On the bluff, obstructing the line-of-sight to Big Eddy, a double reflector and tower was erected to deflect the beam vertically. Fig. 20 shows the aluminum "beam bender" which bends four r-f frequencies between 7,140 and 7,320 mc (megacycles) and Fig. 21 is a profile of the Wasco to Big Eddy path. A diagram of the 4-reflector system layout is shown in Fig. 22. As expected, initial attempts to find the beam failed. In addition to the alignment of the four reflectors the roof-mounted antennas required alignment in two directions. The path alignment procedure

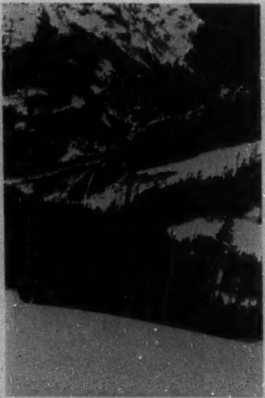


Fig. 18 (left). View from South Rockdale reflector to Rockdale microwave repeater station

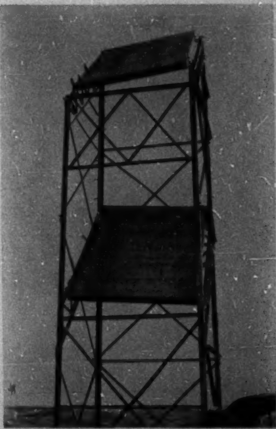


Fig. 20 (right). Big Eddy 11½- by 17-foot double reflector

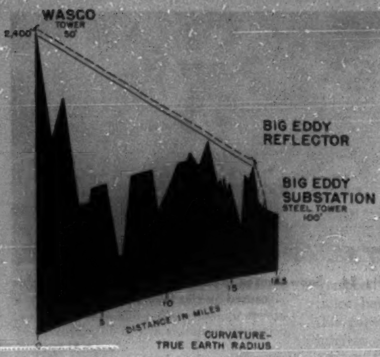


Fig. 21. Profile Wasco to Big Eddy beam path

was simplified by using a temporary source of r-f energy at the reflector site. First the signal was received and peaked at Big Eddy and then the Wasco to Big Eddy path was established. Adjustments were made at each of the six points to obtain maximum signal. Next maximum signal was obtained in both directions by readjusting the two antennas and four plane reflectors for optimum.

Towers and Protection to Microwave Antennas

Most antenna towers on the microwave circuits are self-supporting, designed for a 100-mph (miles per hour) wind and 1/2 inch of ice. A standard tower was developed, with different heights from 54 to 175 feet being obtained by leg extensions. Along the coast, towers are designed for 125-mph wind; one of these towers recently withstood a wind of 158 mph. The Bonneville Microwave System has 33 standard, 8 guyed, 4 special designed, and 3 heavy-duty towers.

In regions of heavy snowfall a different type tower is required. The melting snow creates heavy forces on horizontal members, caused by the upper portions of the snow bank freezing to the structure and as the snow below settles a mass covering a large area is supported by the members. This force acts through a small distance each day but is relentless; it will deform all but the most rugged, special designed structures. A design loading of 2,500 pounds per lineal foot of horizontal member is used where snow depths up to 15 feet are encountered, and 3,000 pounds for above 15 feet. Fig. 23 shows damage to tower members caused by heavy snow.

Where snowfalls 10 to 15 feet are expected, a standard tower is used with a special lower panel having no members smaller than a 4- by 4- by 1/4-inch angle, and very short horizontal members. Where snowfall over 20 feet or extreme icing conditions are expected, a special tower was designed capable of supporting the icing conditions encountered.

Guys, if used, must be at a slope not exceeding two vertical to one horizontal. At a flatter angle 25 feet of snow can break a 5/8-inch high-strength steel guy. All opposing guys on both the North and South Rockdale towers broke the first winter because they were originally installed on a 1 to 1 slope. "Crosby" clips are not suitable. Fig. 24 shows damage caused by slipping guy clips.

Our experience with guyed towers where snowfall exceeds 20 feet has not been favorable. While the tower type is the most economical, it is difficult to hold the reflector or antenna in exact alignment under the severe and uneven stresses produced by snow, which drifts to uneven depths, acting on the guys. An ideal tower would be one with smooth surfaces such as large diameter pipe, with no horizontal bracing below the snow line. The guys on our South Rockdale tower will be replaced in 1957 with supports made of 10-inch pipe.

Towers subject to icing only need not be designed as heavily as for deep snow. There are two types of ice, that caused by sleet, and that formed by frozen cloud crystals. The latter type builds outward in a windward direction in a porous, fluted formation, sometimes to an amazing distance. An extension of 20 feet has been measured at the site of our Mt. Spokane station, on a lookout tower which collapsed. The hard ice caused by sleet causes no difficulties. Towers in locations subject to frozen cloud ice should have no exterior projections such as ladders and airway lighting so the ice is free to slide down the face. Fig. 11 shows an accumulation of ice that has fallen from the tower. Generally, very little ice is formed on the interior. Waveguides and conduit must also be protected from falling ice. While ice will cause no damage to a properly designed tower, means must be provided to prevent ice building out into the beam path, and on the face of large passive reflectors. Also when shields are employed over the antennas, ice cannot be permitted to build up on the underside.

We have been experimenting with vibrators to remove the formation. A vibrator with low amplitude and high frequency is the most desirable. A sudden impact is best for dislodging the ice, and a low amplitude is necessary to prevent imparting vibration or stresses to the waveguides. Results with an electric vibrator were disappointing and it was abandoned in favor of a pneumatic type, which gives a sudden shock.

A $1^{1}/_{2}$-inch pneumatic vibrator mounted on each ice shield, and one at the top of the tower is used with a time switch to energize the vibrators for 15

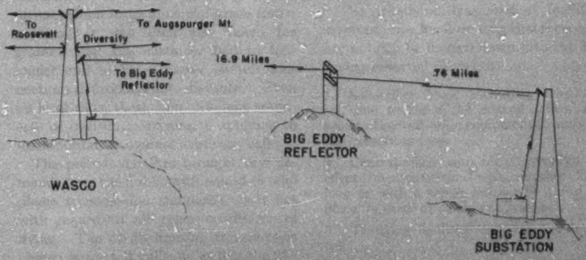


Fig. 22. Wasco to Big Eddy reflector system layout

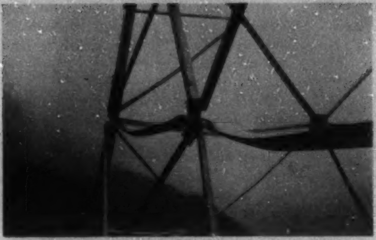


Fig. 23. Damage to steel-tower members caused by deep snow

Fig. 24. Snow damage to tower guy showing rolled thimble caused by clip failure

seconds each 8 hours. A standard service station type 1/4-horsepower air compressor is used to supply air, with an extra storage tank where required to supply sufficient air for 15 seconds without the pressure falling below 60 pounds per square inch. Each vibrator is mounted in an insulated box heated with a 75-watt element to prevent the piston from freezing. The exhaust tube is carried back into the building to minimize condensation. A cartridge-type dehydrator is used on the compressor intake.

This apparatus failed to function properly the first year due to starting difficulties. It has been found that the actuating solenoid valve must be located close to the vibrators or they will not start if there is a long air line between the valve and vibrator. The air must be delivered with an impulse. These vibrators are capable of dislodging the frozen cloud formation. The installation will be altered during the summer of 1957 to overcome the starting difficulties.

Ice shields over antennas must be the shed type to avoid heavy impact from falling ice. The shields seen in Fig. 14 are supported with steel cables, so that vibration will not be transferred from the shield to the structure and the antennas. Ordinary cyclone fabric fencing, supported on a structural steel frame was found satisfactory to protect the horizontally mounted antenna. It is essentially transparent to the microwave radio energy in the frequency band of 7,125 to 7,425 mc.

Conclusions

Reliable communications by microwave can be obtained in mountainous regions with careful design consideration given to the protection of antennas. Covers for parabolic antenna reflectors must be sufficiently strong to withstand the relentless forces of winds. They must be durable to withstand a certain amount of impact and shed snow and ice. Experience has shown that the Thermo-microdome type of antenna cover is the most satisfactory. They are essential for antennas which operate in the 7,145- to 7,425-mc band at certain locations on the Bonneville Power Administration System.

Supporting structures must be carefully designed to withstand winter storms and tremendous loads of deep snows. Structural rigidity is essential to prevent signal loss from antenna misalignment. Antenna reflectors, feedhorns, and transmission lines must be protected from falling ice and snow. Special techniques must be employed to prevent snow from building around and encrusting the antennas.

References

1. MICROWAVE APPLICATIONS TO BONNEVILLE POWER ADMINISTRATION SYSTEM, R. F. Stevens, T. W. Stringfield. *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 1088–99.

2. SOUTHWEST OREGON RADIO SYSTEM FOR BONNEVILLE POWER ADMINISTRATION, D. J. Marihart, D. L. Wylie. *Ibid.*, vol. 75, pt. I, Nov. 1956, pp. 531–38.

3. THE BONNEVILLE POWER ADMINISTRATION LAND-MOBILE COMMUNICATIONS SYSTEM, M. Peckhart, D. E. Johannson. *Transactions*, Professional Group on Vehicular Communications, Institute of Radio Engineers, New York, N. Y., July 1956, p 1.

A Statistical Approach to Telephone Transmission Maintenance

J. B. PRINGLE
NONMEMBER AIEE

G. GAUDET
MEMBER AIEE

Synopsis: Trunk-net-loss deviation, i.e., the difference between measured and expected value of trunk loss, may be considered to be the outcome of a complex process, the result of many variables. In general, a series of such observations forms a statistical distribution. Such distributions have variability, dispersion, or spread. The measure of this is standard deviation, designated as σ'.

This variability can be analyzed into two components, one due to findable causes, and one due to chance. If action is taken when chance alone is operating, more variability is introduced than was present previously. If action is not taken when findable causes are operating, unnecessary troubles are allowed to vitiate the process.

The Shewhart control chart is a convenient device for detecting findable causes on background of chance causes. Its application to the telephone transmission maintenance problem is discussed.

A COMMERCIAL telephone system must of course be designed, maintained, and operated so as to provide a grade of communication service satisfactory to customers. Good telephone communication service involves the following:

1. Ease of obtaining the desired connection.
2. Once the desired connection has been obtained, ease of conversing without effort and with an auditory feeling of naturalness.
3. Freedom of interference from noise or crosstalk.
4. Safety to persons and physical property, both for customers, company, and employees.
5. Reasonably low cost of service consistent with sound economy.

In actuality, these five factors are closely interrelated. It is one of the functions of engineers and management to assess the relative importance of each requirement in allocating expenditures calculated to provide best service.

Requirements 1, 3, 4, and 5 are all basic, but in final analysis, they are only auxiliary conditions to permit fulfilment of the main objective of the system, which is to make possible easy and natural conversation between any two customers, irrespective of station locations and distance between them. For purposes of this paper, requirement 2 is of principal interest.

The ability to carry out effortless and natural conversation, i.e., exchange of intelligence, can be dissected so as to identify two fundamental requirements:

1. Volume, i.e., loudness of reception.
2. Quality, i.e., naturalness and fidelity in reproduction of tone, accent, and articulation.

With modern techniques and instrumentalities, quality of telephone transmission is no longer a critical problem. Telephone instruments and facilities in common use are all adequate to provide satisfactory quality performance over a speech frequency band of roughly 200 to 3,200 cycles per second which wide experience confirms as adequate.

The foregoing elimination leaves volume as the troublesome factor, which presents many design problems, and poses equally challenging problems from the transmission maintenance standpoint because of the inherent tendency of transmission volume levels to vary over wide limits, as discussed later.

Volume Standards, Past and Present

In earlier stages of telephone development, there has been a widely accepted "standard" of 31 db (decibels) as an ultimate limit for commercial-grade transmission. It was then also taken for granted that the customer had to talk louder and to listen closer on toll connections in proportion to distance. Even on local calls, there were different standards of over-all transmission applying to metropolitan as against suburban calls.

The present day has brought new demands from the public with regard to telephone transmission standards, as it has with regard to all other conditions of living. The 31-db limiting transmission objective for toll calls, as well as the 24-, 22-, and 20-db standards for intersuburban, suburban metropolitan, and intrametropolitan are now as obsolete as the model "T" Ford. Surveys of customers' opinion and other investigations indicate a public expectation that telephone transmission, irrespective of distance, should be about as good as person-to-person air transmission across a desk in an average quiet office. In technical terms, this corresponds to an over-all effective loss of some 6 db down to 0 db, quite a contrast against the old 31-db toll and 24-, 22-, and 20-db exchange standards.

Another challenging factor has arisen in the last decade. A direct-distance dialing plan, with the direct dialing performed by the operator, or in a later stage of progress by the customer, has been designed to increase speed of connection on long-distance calls, and over the years to accomplish economies in operation. From the transmission engineer's viewpoint, this new development entails added difficulties due to the possible introduction of a greater number of links in the longest connections. This in turn necessitates, in order to meet the new over-all low transmission-loss objectives, a more than ever stringent design of transmission facilities to keep objective transmission losses of each link to a minimum, with greater and more efficient effort than ever to control the spread of loss variations per link so that their cumulative effect will not have any disastrous effect on the over-all transmission performance.

The design of transmission facilities to meet modern requirements keeps in step with the latest advancements in the fields of electronics and industrial techniques in general. The main subject of this paper has to do with the equally important problem of controlling variations or deviations in performance of facilities, assumed to have been properly designed to meet modern objectives, after they have been placed in service.

Transmission Maintenance Methods

The telephone transmission facilities involved in a local or long-distance connection can be broken down into the following essential elements:

1. Subscriber loop, including telephone set, line to serve the central office, and essential central-office equipment to activate the subscriber loop.
2. Central-office equipment required to effect interconnection with another local loop or with a trunk leading to a distant office, in same or remote exchange.
3. Trunking facilities, interlocal or intertoll.

In modern practice, items 1 and 2 are

Paper 57-784, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted March 25, 1957; made available for printing May 20, 1957.

J. B. Pringle and G. Gaudet are with The Bell Telephone Company of Canada, Montreal, Que., Canada.

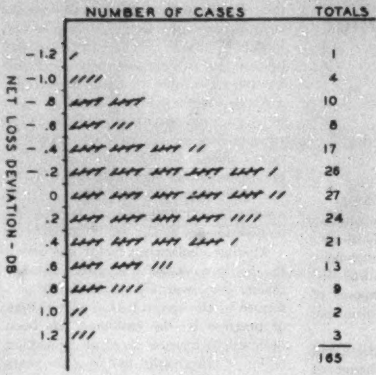


Fig. 1 (left). Frequency distribution

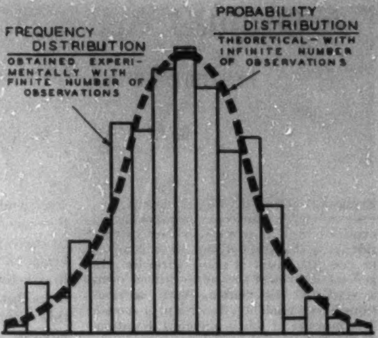


Fig. 2 (right). Probability distribution

fairly fixed in design and not normally subject to troublesome transmission-loss fluctuations. Item 3 is the main object of the maintenance methods based on statistical methods presented in this paper.

The control of the nominal trunk loss is a function of design; the control of trunk-loss variations (including detection of complete failures) is a function of transmission maintenance.

Transmission maintenance is considered frequently in terms of two different approaches: preventive, and corrective maintenance. The first consists of doing work in anticipation of failure, and the second consists of correcting troubles only after a failure. These concepts rather cut across the economics of the matter; it is distressingly easy to do redundant work in expectation of failure that never occurs, and it is also easy, if one adheres rigidly to the corrective idea, to let maintenance deteriorate to a point where minimum working standards are not met. It appears that what is wanted is some method of directing maintenance effort toward and only toward profitable ends, i.e., toward the reduction of actual and potential trouble conditions without wasting it in directions where no trouble is indicated. This points up the desirability of a suitable detection device for probable trouble conditions. A device of this nature is available in the Shewhart control chart.

The control chart is a statistical decision-making tool that has been used to great advantage in quality control work on manufactured products. It is essentially a mathematical or statistical device, using measurement data as raw material. It can be applied in any medium where measurements can be obtained. There is a tendency to think of it as applicable to a "nuts-and-bolts" type of product only, but, if suitable measurements are available, it can be applied to any kind of "outcome-of-a-process," e.g., trunk-net-loss deviation.

Variability and Probability

Consider the volume obtained over a telephone connection to be the outcome-of-a-process. It is an extremely complex process, dependent on not only the basic design parameters of the system, but also on a host of factors tending to deviate the actual individual result from the design expectancy. Some of these factors may be:

1. Temperature and weather effects.
2. Manufacturing deficiencies.
3. Human errors in adjustments and measurements.
4. Deviations in precision of measuring apparatus.
5. Errors in construction, such as missing loading coils, wrong type of coils, wrong coil spacings, etc.
6. Physical faults developed since original construction, such as high-resistance joints, faulty insulation, leaky condensers, etc.
7. Pure chance causes such as cumulative combinations of individually tolerable deviations.

The foregoing list could be extended indefinitely.

The measurement considered here is the trunk-net-loss deviation, or the departure from design trunk loss that is experienced due to the causes, outlined in the foregoing, that are introduced by the process. This measurement is subject to variability; i.e., on a number of measurements made from time to time, on the same or different facilities, the same results may not always be obtained. The results will show some scatter, i.e., they will form a distribution of some kind. The spread of this distribution is a measure of variability.

It is the basic statistical problem to resolve this variability into two components: 1. that caused by chance, or by a great many small independent factors which cannot profitably be analyzed, and 2. that caused by significant factors, "assignable causes," or causes which can be found and removed from the process.

It is well known to the statistician that if action is taken when chance variability alone is present, more disorder is created than if nothing had been done. Conversely, if no action is taken when findable causes are present, this allows unnecessary disorder to exist in the process. Statistical methods, such as the control chart, provide criteria for rational action in the face of the variability bound to be experienced.

The raw material used in the preparation of control charts are observational data, the results of actual measurements of whatever characteristics are concerned. In the present instance trunk-net-loss deviation, in decibels, is used as the measurement. This is the difference between the actual measured trunk loss and its design loss. Under ideal conditions this would be zero, but, due to the operation of chance causes or assignable causes, it may be expected, in a large number of observations, that a distribution of these values will be found spreading around zero in some fashion, as indicated by the frequency distribution in Fig. 1.

The results tend to center around an average value, with the frequency of occurrences diminishing in each direction

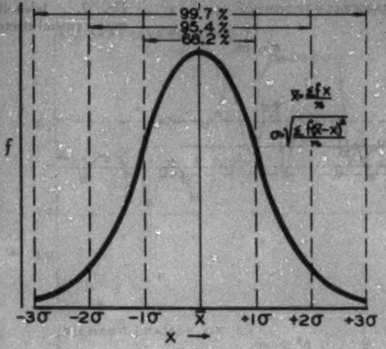


Fig. 3 (left). Normal distribution

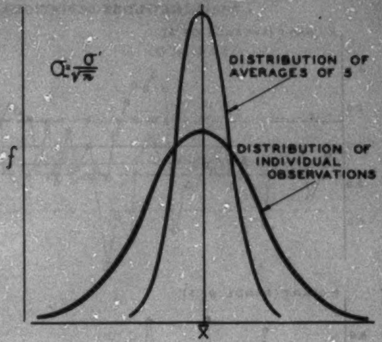


Fig. 5 (right). Distribution of averages of sample

from the average; i.e., in rough form, a normal distribution. It can be shown that this is the form of distribution obtained when a great number of small independent causes are working to change things from a norm, or central value, indicated by the average.

The normal probability distribution is a convenient mathematical model for the frequency distributions obtained in practice.

Probability may be defined as the ratio of "looked for" outcomes of a process to the total outcomes, expressed as a percentage. If all the outcomes of a process form a distribution, as they do in a frequency distribution, or its idealized version (the normal probability distribution), the total area under the curve will represent 100% probability. In other words it can be taken as certain (if the process does not change) that a subsequent observation from the process will fall somewhere within the distribution. Moreover, the probability of an occurrence having a value falling between any two ordinates of the curve is the ratio of the area under the curve between these ordinates to the total area under the curve.

Fig. 2 indicates the relation between the practical frequency distribution which is obtained experimentally with a finite number of observations and the theoretical probability distribution which, from the definition of probability, can really be obtained only with an infinite number of observations, but which can be readily estimated from a finite number of observations, i.e., from a frequency distribution. If a histogram is made, such as indicated by this frequency distribution, with only a few observations, it will be a poor approximation of the probability distribution. The fit will be rather ragged. With many observations, however, the fit is likely to be good.

The normal distribution can be described by using only two figures: 1. a measure of central tendency, the mean or average designated in Fig. 3 as $\overline{X}$; and 2. a measure of the spread of the distribution. For this purpose the root mean square of the individual deviations from the mean is a convenient measure. This is known as the standard deviation, and is designated by σ. When a distribution of individual observations is being dealt with, the prime notation is usually added, as σ'.

In a normal distribution 68.2% of observations will fall within $\pm 1\ \sigma$ of the mean, 95.4% within $\pm 2\ \sigma$, and 99.7% within $\pm 3\ \sigma$. In other words, there is a 99.7% probability, which would be close to certainty, that an observation in the group being considered, due to the operations of chance, will fall within $\pm 3\ \sigma$ of the mean.

The Control Chart

The basic idea behind the control chart is extremely simple. Fig. 4 shows the probability distribution of a variable subject to chance (the shaded distribution), and the three σ limits thereof extended horizontally. In practice, this is obtained from the frequency distribution of a set of "base-period" observations on the process. Now if from this distribution

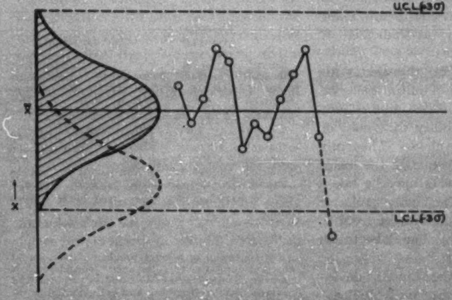


Fig. 4. Control chart for individual observations

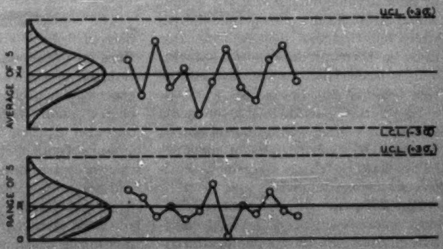


Fig. 6. Control chart, X̄ and R charts

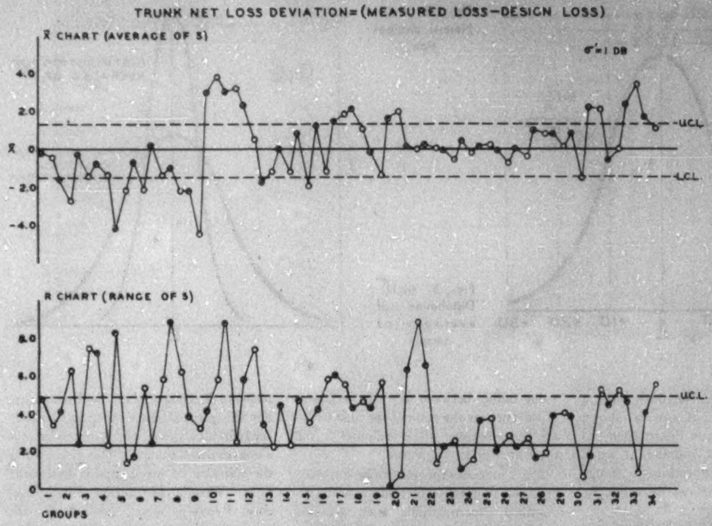


Fig. 7. Intertoll trunks, control charts

or rather from the process generating it, a series of individual measurements is taken, the values obtained will plot in some such form as illustrated. They will tend to scatter around the average, most values being close to the average, but practically none occurring outside the three σ limits. The sequence in which the values will fall will be completely unpredictible.

Supposing, however, some cause affecting all observations is introduced into the process. This is what is called an assignable cause. All subsequent observations would be affected by it, in fact one would be dealing with the new probability distribution which might be represented by the dashed curve in Fig. 4. Now it is obvious that if continuous observation is made of individuals from the process under these circumstances, sooner or later a value outside the ±3 σ limit would be obtained. This is the criterion which tells that an assignable or findable cause has been introduced into the process and indicates that appropriate corrective measures should be taken.

In practice individual measurements are seldom plotted on the control charts. Averages of small samples of measurements make the charts much more sensitive as a detection device.

Fig. 5 illustrates a statistical relationship concerning this. The broad distribution represents the probability distribution of a measurement on the individual items of product of a particular process. If samples are taken from this distribution in groups of five, and the averages of each of these groups are computed, these averages would also form a normal distribution, but one with considerably smaller spread. The standard deviation of the distribution of averages of five so obtained would be equal to the standard deviation of the distribution of individual observations divided by the square root of five, as illustrated in the narrow distribution.

In practice measurements are usually taken from the product in groups of five, and not only the averages of these, but their ranges as well, are plotted. The range is the difference between the maximum and minimum values obtained in the sample of five. Two charts are maintained simultaneously, as indicated in Fig. 6.

The control limits on the charts are derived from a series of initial observations on the process, which give a good estimate of the variability that may be expected due to chance alone. Then, any observed point falling outside the limits on either of these charts may be taken as an indication of the highly probable existence of a findable cause of trouble, and appropriate action should be taken when this occurs.

A good estimate of the standard deviation of the distribution of individual observations can be obtained conveniently from the average of the ranges of samples of five, and from this the control limits for the charts can be calculated. Constants for computing the control limits may be obtained from any standard text on statistical quality control. A number of these are listed as references. However, for convenience, a brief table indicating how control limits are calculated is given in Table I.

It is essential that the samples of five be taken as "rational subgroups," i.e., there should be a basic similarity in the units constituting the sample, e.g., they should be trunks with practically the same make-

Table I. Factors for Control Charts

	Chart for Averages*	Chart for Ranges†		
No. of Observations in Sample	Factor for Control Limits, A	Factor for Central Line, d_2	Factors for Control Limits D_1	D_2
2	2.121	1.128	0	3.686
3	1.732	1.693	0	4.358
4	1.500	2.059	0	4.698
5	1.342	2.326	0	4.918

* $\bar{X}$; central line, $\bar{X}'$; upper and lower control limits, $\bar{X}' \pm A\sigma'$.

† R; control line, $d_2\sigma'$; upper and lower control limits, $D_1\sigma'$ and $D_2\sigma'$.

$\bar{X}'$ is the mean of a large number of base-period observations. In the instances considered, it may be assumed to be zero. σ' is the standard deviation of the distribution of individual observations. It may be obtained from a large number of base-period observations.

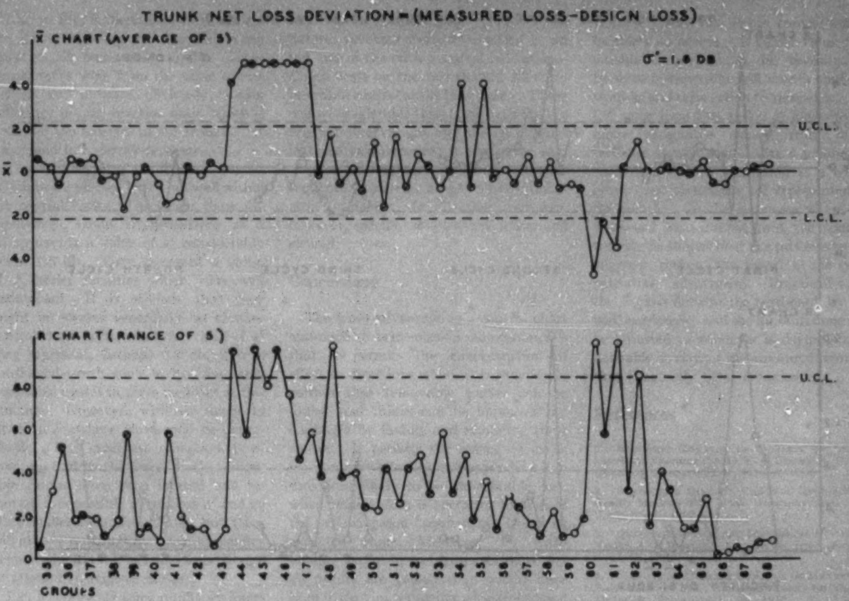


Fig. 8. Intertoll trunks, control charts

up, terminating in the same offices, etc. If an assignable cause is present which affects all units in the sample, its presence will be much more easily detected as an "out-of-control" point if the sample constitutes a rational subgroup.

Procedure

The specific purpose, then, is to control trunk-net-loss deviation in the plant by the use of $\overline{X}$ and R charts. Trunk-net-loss deviation is the difference between the measured 1,000-cycle loss of the trunk and its design loss, the latter appearing on the circuit-layout record card. These deviations may be positive or negative, but will, in general, form a distribution which should center around zero.

In any given office, the basic parameters of this distribution can be estimated from a large sample of observations. This is done at present by using a stroke chart (frequency histogram); a figure for standard deviation is obtained from this by fitting a template to the envelope of the distribution.

It is proposed that sample observations be taken in groups of four or five similar trunks, e.g., from the same carrier system, similar systems, coterminal facilities, of same make-up, (rational subgroups), etc., and X-bar and R charts for the averages and ranges of these samples be plotted. Control limits can be obtained from the basic parameters $\overline{X}'=0$ and standard deviation σ' obtained, at least initially, from the over-all distribution as indicated in the foregoing.

Note that in this application it is suggested that σ' be obtained from the distribution of individual observations. At a later stage, if within-subgroup variability becomes large with respect to between-subgroup variability, σ' may be computed from the average of a large number of measured ranges, using the relation

$$\sigma'=\frac{\bar{R}}{d_2} \qquad (\bar{R}=\text{average range})$$

As the worst assignable causes are found and removed, it is entirely possible that it may be found desirable to employ, to a degree, different charts for different kinds of facilities because it appears likely that the inherent chance variability may differ considerably between types of facilities.

In any event, after a period of operation during which assignable causes will have been found and removed, a reassessment of σ' should be made which will result in narrower limits on the control chart.

Initially, it is thought that measurements should be made in both directions, resulting in two samples of five from a group of five trunks. For administrative purposes, later, it might be desirable to make the observations in one direction only, at a time.

The observations should be made, recorded, and plotted right at the testboard. If action is called for, it should be undertaken with a minimum of delay.

If the $\overline{X}$ and R values obtained for a rational subgroup fall within control limits, no action should be taken. If either or both the $\overline{X}$ and R values for a rational subgroup fall outside the control limits, investigation should be undertaken to find an assignable cause for the condition, to lead to the removal of this cause.

It is expected that many assignable causes can readily be detected and removed. If $\overline{X}$ is out of limits, a cause common to the group is generally indicated, and the trouble should be looked

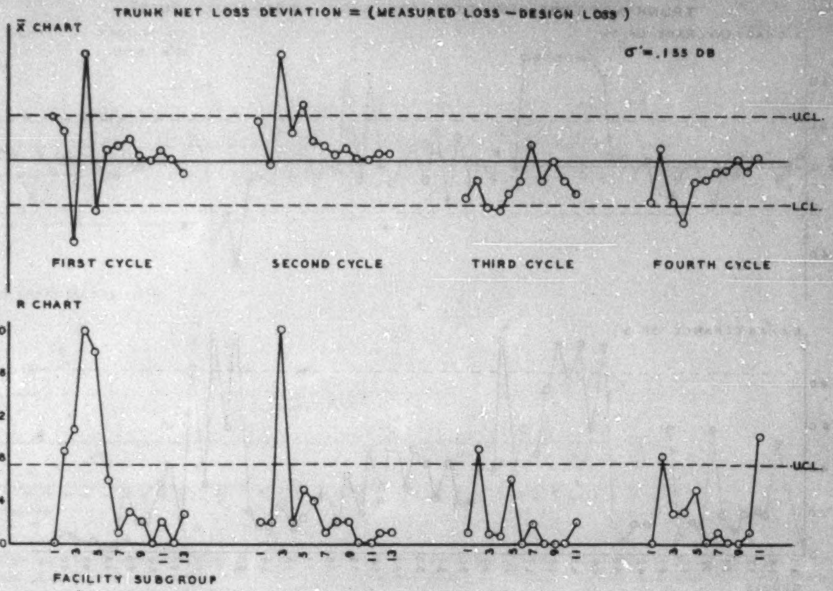


Fig. 9. Intertoll trunks, control charts

for in something involving the whole group, such as a carrier repeater. If a range outage is found, it will generally indicate a cause involving only the one facility. In either case, comparisons of the measurements in the two directions may be valuable in locating trouble.

The amount of sampling to be done is not a critical matter. All facilities from a given office can be covered in a certain period if required, or only occasional random samples, or samples from suspected trouble areas, may be covered. To a great extent, the amount of work done can be adjusted to whatever may be justified by found trouble.

It is important that action be taken to remove causes, rather than to introduce compensating adjustments. The control chart will only indicate the very probable presence of a findable cause. A real effort should be made to find it before any adjustment is made.

Perhaps one of the most frequent and baffling trouble causes may be found to be simply maladjustment of gain in the first place. In these instances it is desirable to proceed with caution before readjusting. Probably the best procedure would be to take a rational subgroup of observations from this trunk itself, in time, and readjust so that the average so obtained would become zero. A quicker but less precise method that has been suggested would be to adjust halfway to zero without taking additional measurements. Undoubtedly a suitable procedure can be evolved. The present procedure of readjusting to zero on the basis of deviations that are small compared with $\pm 3\ \sigma'$, it can be shown, tends to increase over-all variability.

There will be instances where adequate investigation may be protracted, but it is expected that procedures can be developed that will ensure maximum probability of determining what the findable causes are with a minimum of effort. Simple experimental designs, which can be put in the form of definite procedures for the testboard man to follow, would help in this respect.

Some Results

Fig. 7 shows some actual results from a certain office. Averages and ranges from rational subgroups of five trunks each are plotted. Control limits were based on a value of $\sigma' = 1$ db, which was taken arbitrarily, but which is considered to represent fairly good maintenance at present, and the office involved was thought to be in about this class.

The results show that a great many groups of trunks tested showed out-of-control conditions on the basis of $\sigma' = 1$. Man power was simply not available to investigate properly all these cases. It was necessary to approach the problem realistically by computing a value of σ' from the actual observations, and plotting chart control limits on this basis. Such a chart is shown in Fig. 8. It represents a value of σ' of 1.6 db.

With the wider limits, there are considerably fewer out-of-control indications in the new chart. The outstanding cases register, but they are not so many that they cannot be properly attended to. Generally, too, it is desirable that these outstanding cases be corrected before much effort is expended on correcting smaller deviations.

The first notable group of out-of-control points represented a group of facilities that were known to have certain deficiencies. These were cleared up in the course of the following month, and are now among the best of that type.

Also in Fig. 8 there are two points on the $\bar{X}$ chart showing out-of-control in one direction of transmission only. These two samples were from the same carrier system, and indicate, obviously, a gain deficiency in one direction only, which is a clue to the location of the trouble. It was found in a carrier repeater.

Toward the end of the $\bar{X}$ chart a series of points are plotted which fall well within the control limits. Obviously these, by themselves, could be considered as a group having a value of σ' considerably below 1.6 db. They represent a group of *K*-carrier facilities which were well maintained. It is obvious that they might be treated separately on another control chart, using whatever value of σ' they represent, perhaps 0.8 db, if it is considered worth while to look for small assignable causes in these facilities at this juncture. However, with so much to clean up elsewhere, this would be unprofitable. This gives an indication, however, that when the worst of the assignable causes have been located and removed, it is possible to reassess σ' and go after smaller assignable causes; but, there will always remain some basic variability beyond which it will be found unprofitable to attempt to remove causes.

Fig. 9 represents some results on interlocal trunks. Samples of five were taken from 13 groups of similar trunks radiating from a certain office. A first sample of five from each of these groups constitutes the *first cycle*, a second sample of five, the *second cycle*, etc.

The standard deviation of the individual observations σ' was calculated from the whole group of observations. A number of out-of-control points were indicated, as shown in the figure. Subsequent investigation, of a protracted nature, revealed definite troubles in all but two of the out-of-control indications. Check tests on the two showed them to be within control at the later date. There was a considerable time lapse between the original observations and the check, and in these two instances it cannot be said whether it was some trouble that subsequently cleared, or an observational error originally. In all other instances, however, specific troubles were found and cleared.

Conclusions

The main advantage in a control-chart approach to transmission maintenance is that it permits the concentration of effort in directions where it is practically certain that removable causes can be found, and things can be improved accordingly by finding and removing these causes. It permits the saving of much maintenance effort by not expending it in directions likely to be unprofitable, i.e., when chance alone is operating to deviate the transmission equivalent from its expected value. Moreover, by withholding action in such cases, better overall results are gained than would have been if the compensating adjustment had been resorted to.

It might be well, in conclusion, to emphasize the simplicity of application of the method. Given a certain amount of base-period data, the engineer can compute control limits and set up average and range charts for any group of facilities, such as those controlled by a given office. The testboard man can select rational subgroups of four or five, measure them, and plot the average and range of each sample. If these values fall within control limits, he takes no action. If either value falls outside control limits, he investigates. In most instances he will readily find the trouble, and take action to remove it. At the same time this action is likely to remove assignable causes from other facilities as well, since quite a number of such causes will be common to a whole group, not necessarily all represented in the sample. In the few cases where the testboard man cannot find the specific trouble, he should refer the problem to the engineer, rather than resort to the compensating adjustment. Practically all the work is done at the testboard, by the testboard man; and the amount done can be adjusted to whatever is deemed to be desirable in view of the amount of trouble that is actually being experienced.

References

1. Economic Control of Quality of Manufactured Product (book), **W. A. Shewhart.** D. Van Nostrand Company, Inc., Princeton, N. J., 1931.

2. Statistical Quality Control (book), **E. L. Grant.** McGraw-Hill Book Company, Inc., New York, N. Y., 1952.

3. Quality Control and Industrial Statistics (book), **A. J. Duncan.** Richard D. Irwin, Homewood, Ill., 1952.

4. Manual on Quality Control of Materials. *Publication no. 15-C*, American Society for Testing Materials, Philadelphia, Pa., 1951.

5. Control Chart Method of Controlling Quality During Production. *War Standard Z1-3*, American Standards Association, New York, N. Y., 1942.

6. Interpretation of Engineering Data: Some Observations, **H. F. Dodge.** *Marburg lecture*, American Society for Testing Materials, Philadelphia, Pa.; also available as *Monograph 2375*, Bell Telephone Laboratories, New York, N. Y., 1954.

7. Facts From Figures (book), **M. J. Moroney.** Penguin Books, Ltd., Harmondsworth, England, 1956.

8. Statistical Methods in Quality Control. *AIEE Special Publication*, Jan. 1948.

◆

Discussion

G. D. Edwards (Consultant Engineer, Cape Elizabeth, Me.): The paper discusses, in admirably simple and direct fashion, the application of fundamental control principles to a problem involving a considerable number of groups of causes—there often being very little correlation among the groups, but frequently considerable correlation among the causes within individual groups. It cites effective results from use of the procedures described in specific instances, and shows reason for anticipating even greater effectiveness as further experience permits refinements in approach and method.

The authors make a striking point in urging avoidance of attempts at compensating adjustments until completion of every reasonable effort to eliminate findable causes. Failure to heed this advice of course leads to multiplication of the number of causes, and their individual identification and elimination become increasingly difficult.

H. E. Corey, Jr. (American Telephone and Telegraph Company, New York, N. Y.): The paper gives well-deserved recognition to the importance of the statistical approach in the design of a modern telephone maintenance program. The control chart may prove to be a very effective tool for reducing trunk-loss variation and there is certain to be widespread interest in the outcome of future trials of the plan described.

While the authors have discussed only the application of the control chart there are, of course, a number of statistical techniques which might contribute to an effective and economical solution to the trunk maintenance problem. At this stage it might be very profitable to explore some of these other avenues as well, with an eye to developing a maintenance procedure, founded on sound statistical principles, requiring a minimum of measurement, analysis, and judgement. Another consideration to be kept in mind in this era of automation is the matter of possible future reduction of such testing and decision-making procedures to simple machine operations.

All of this leads up to a few thoughts about a possible procedure using simple maintenance criteria based on a knowledge of the probable behavior of various types of trunk facilities with time. If enough could be learned about this behavior from carefully planned field evaluations, it might be found practicable to write maintenance instructions which fundamentally look approximately like the following:

1. Never make adjustments in trunk loss, except after a repair operation, or 2. Adjust the loss of each trunk to some nominal value at the end of predetermined uniform time intervals, and in addition, 3. Hunt trouble only when the difference between sequential measurements on a given trunk indicates that a certain time rate of change in loss has been exceeded.

Note that if instructions 1 and 3 are applicable, this says, "Don't do anything unless sequential measurements indicate trouble." Then it is a simple "go," "no-go" proposition and it is hard to conceive of anything less complicated than this—or incidently, of anything much less difficult to mechanize.

Now, this procedure may not be feasible. In fact it may not work at all. But the attainment of such simple objectives is one which is loaded with challenge, both for the maintenance engineer and for the statistician.

On a different subject for a moment, in the paper are listed five requirements for good telephone service. Requirement 2 is divorced from the other four and labeled the principal one for the purpose of the paper. The logic for doing this is quite clear, and this separation proves to be quite acceptable as the paper develops. There is very strong interaction between requirements 2 and 3, however, and from the customers' point of view it is not possible to separate ease, effort, and naturalness from conditions of noise and crosstalk. Of course, this does not prevent these problems from being tackled separately, but one should constantly be reminded that both are vital to customer satisfaction.

Finally, referring to the difference between measured trunk loss and design loss this statement is made: "These deviations may be positive or negative, but will, in general, form a distribution which should center around zero." This statement is perfectly true as it stands, but could be misunderstood. Note carefully, the authors say "should center" and not "do center." In practice they almost always center about some value different from zero and too large to be neglected. It is clearly the objective of the control chart or any other statistically derived procedure to control this deviation average, termed bias, as well as the spread among individual deviations.

Safety Systems for Nuclear Power Reactors

I. M. JACOBS
ASSOCIATE MEMBER AIEE

THE power output of a conventional fossil-fired power plant is limited by the rate at which fuel can be fed to the boiler. No such convenient limitation is available for a nuclear reactor, which theoretically has no upper limit of power level. After allowance is made for all restrictive negative coefficients of reactivity, the maximum power is still excessive. Clearly, some artificial means of limiting the power is indicated.

On research and plutonium-producing reactors, the problem has been solved. Ionization chambers, amplifiers, and controllers are arranged to measure the power level and trip (or scram) the reactor by inserting safety rods if the power level is excessive. That the solution is satisfactory is evidenced by the record of safe operation of dozens of reactors for over a decade.

A nuclear power station must maintain this superb record of safety and at the same time establish new standards for continuity of service. In short, the safety system must never fail to scram the reactor on a bona fide power excursion; and it must almost never scram the reactor for any other cause, including component failure. These two objectives, if pursued to the ultimate, are diametrically opposed, and safety considerations must always take first place. However, great improvements in continuity of operation can be achieved without sacrifice of safety. This paper is a comparative analysis of the several safety system arrangements viewed from both a mathematical and operational standpoint.

Definition of Terms

The following terms, as defined, are basic to the analysis presented in this paper:

Safety channel—an ionization chamber, amplifier, and controller system designed to scram a reactor on excessive flux (or power level).

Fail safe safety channel—a safety channel[1] that scrams if it develops a fault that would keep it from tripping on excessive flux. Note that the term "fail safe" is an objective of good design and can never be completely achieved in reality.

Operational fault—a fault which prevents a safety channel from scramming on excessive flux but which is not revealed until the controller is tested. (Note that this is the fault which prevents fail safe from being a reality.)

Spurious fault—a fault which results in a scram when the flux is not excessive.

Safety System Arrangements

SINGLE SAFETY CHANNEL

One safety channel can be arranged as shown in Fig. 1. The ionization chamber monitors the reactor flux. If the flux is excessive, the controller contact opens, the safety rod holding device is deenergized, and the rod enters the reactor, causing a scram. The reactor is shown in the scrammed condition.

In the case of a single safety channel, each spurious fault initiates a false scram. Thus, the rate of false scrams due to spurious faults is

$$R_s = F_s \tag{1}$$

where R_s is the rate of false scrams due to spurious faults, and F_s is the average rate of spurious faults.

If an operational fault should occur, the reactor would be without protection until the next testing. The time the reactor is unprotected may be found as follows:

F_0 = average rate of operational faults
F_0dt = probability that an operational fault will occur in time dt
τ = interval between testings
$(\tau - t)$ = time the fault which occurred at time t will remain undetected[1]
t = time a fault occurs after beginning of test period

It can be shown that the following integral is an expression for the total time T the process is unprotected between testings

$$T = \int_0^\tau F_0(\tau - t)dt \tag{2}$$

which becomes

$$T = \frac{F_0\tau^2}{2} \tag{3}$$

F_0 takes on a limiting maximum value of $1/\tau$ because only the first operational fault during any one test interval has any significance. A better comparative index of systems can be determined by making

---

Paper 57-906, recommended by the AIEE Nucleonic and Radiation Instruments Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Pasco, Wash., August 28–30, 1957. Manuscript submitted March 25, 1957; made available for printing July 24, 1957.

I. M. JACOBS is with the General Electric Company, San Jose, Calif.

The author is indebted to D. B. Klapper of the Commonwealth Edison Company for assistance in checking the mathematical derivations in this paper.

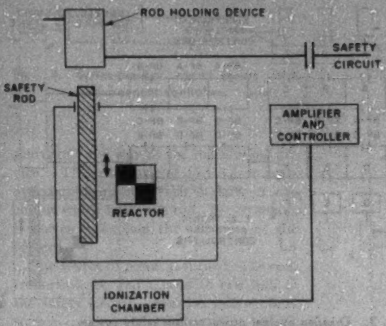


Fig. 1. A single-channel safety system

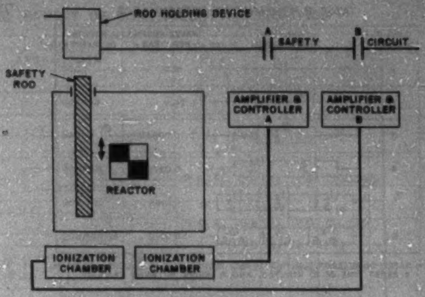


Fig. 2. Two-channel series safety system

use of a figure of merit defined as follows:

$$\text{FM (figure of merit)} = \frac{\text{total time between testings}}{\text{time unprotected between testings}}$$

or

$$\text{FM} = \frac{\tau}{T} \tag{4}$$

For the single safety channel

$$\text{FM} = \frac{2}{F_0 \tau} \tag{5}$$

Two-Channel Series Safety System

An improvement in the safety or FM of the safety system can be achieved by connecting the output contacts of two independent channels in series in the safety circuit; see Fig. 2. With this arrangement both channels must suffer a coincidental operational failure before the process is unprotected. However, since either channel can initiate a scram, the rate of false scrams from spurious faults will double.

The rate of false scrams for this series circuit is

$$R_s = 2F_s \tag{6}$$

and the

$$\text{FM} = \frac{3}{F_0^2 \tau^2} \tag{7}$$

(A typical derivation of this nature is presented in the appendix.)

If F_0 and τ are kept small, the FM of this circuit increases many fold over the single channel safety system.

Combination Series and Parallel Safety Systems

The foregoing analysis confirms the intuitive feeling that connecting the output contacts of additional independent safety channels in series improves safety at the expense of an increased rate of false scrams. Since the objective is to reduce false scrams (without sacrificing safety), one is led to try parallel arrangements of contacts.

In the parallel arrangements, spurious faults must occur in coincidence in order to cause a false scram. A spurious fault will cause one controller to trip and remain tripped. If another controller trips from a spurious fault before the first one is repaired or reset, the reactor is scrammed. Thus, a new variable, the time to reset or repair a controller, is introduced. This is called t_R.

Table I lists the basic circuits. Circuits 1 and 2 are repeated for reference. Circuit 3 is composed of two controller contacts in parallel. Since controllers of a parallel arrangement must suffer a coincident spurious fault within the reset time, the rate of false scrams can be kept low. However, an operational fault on either channel will prevent a scram; the FM is only one half as high as the value for the single controller.

The series-parallel arrangement of circuit 4 offers real promise. Since only certain combinations of faults can cause malfunction, the rate of false scrams is reduced and the FM remains at a high level. This series-parallel circuit approaches the FM performance of the series circuit and the low rate of false scrams of the parallel circuit.

Circuit 5 is an arrangement of contacts designed to require coincident trips from any two of the three controllers to initiate a scram. The rate of false scrams is slightly reduced and the FM is reduced; but this circuit compares favorably with

TABLE I PERFORMANCE EQUATIONS

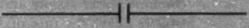
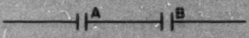
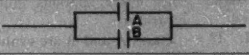
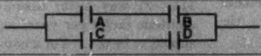
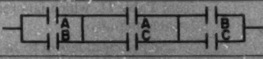
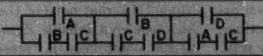
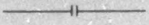
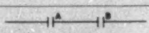
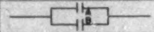
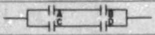
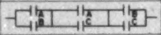
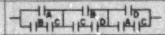
CIRCUIT NUMBER	CIRCUIT ARRANGEMENT	RATE OF FALSE SCRAMS	FIGURE OF MERIT
1		F_s	$\frac{2}{F_0 \tau}$
2		$2F_s$	$\frac{3}{F_0^2 \tau^2}$
3		$2F_s^2 t_R$	$\frac{1}{F_0 \tau}$
4		$8F_s^2 t_R$	$\frac{3}{2F_0^2 \tau^2}$
5		$6F_s^2 t_R$	$\frac{1}{F_0^2 \tau^2}$
6		$12F_s^2 t_R$	$\frac{1}{2F_0^3 \tau^3}$

TABLE II COMPARATIVE PERFORMANCE

CIRCUIT NUMBER	CIRCUIT ARRANGEMENT	FALSE SCRAMS PER YEAR *	FIGURE OF MERIT *
1		12.0	1,096
2		24.0	901,500
3		0.00548	548
4		0.0279	450,750
5		0.0164	300,500
6		0.0323	82,400,000

* BASED ON 12 SPURIOUS FAULTS AND 2 OPERATIONAL FAULTS PER YEAR, A RESET TIME OF 10 MINUTES, AND 8 HOURS BETWEEN TESTING.

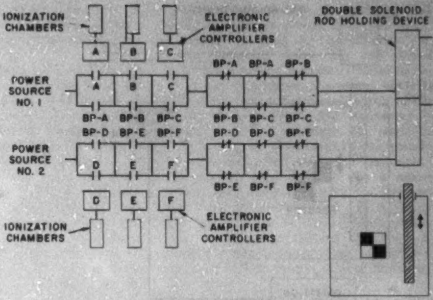


Fig. 3. Dresden nuclear power station safety system

circuit 4 with a saving of one safety channel.

Circuit 6 is an arrangement of contacts designed to require coincident trips from any two of the four controllers to initiate a scram. The rate of false scrams is somewhat higher than most of the other circuits, but the cubed terms in the denominator of the expression of the FM point the way to an exceedingly high safety performance.

Comparative Performance

An absolute evaluation of performance of a system is no better than the data that goes into the calculations. Performance data are difficult to obtain because the failure rates are so low as to make record keeping uninteresting. Some new insight can be gained as to the comparative performance of the various circuits by assuming plausible values for failure rates, for reset times, and for intervals between testings.

Assume:

$F_s = 12$ spurious faults per year (per controller)
$F_0 = 2$ operational faults per year (per controller)
$t_r = 10$ minutes' reset time
$\tau = 8$ hours between testings (once per shift)

No claim is made that this is an excellent, poor, or average quality controller, only that the values are plausible.

The assumed time values must be expressed in years or other consistent units and the equations of Table I evaluated.

The results are shown in Table II. If continuity of operation is a requirement second only to safety, the case for the series-parallel circuits is clear.

Interaction Between Controllers

In the analysis thus far it has been assumed that the faults are random in time; i.e., that they are not forced in any way. Spurious faults may follow a pattern which is far from random. Due to the nature of these controller systems, the low levels, and wide dynamic ranges over which they operate, they are inherently subject to line transients and induced noise. The disturbing influence which causes one controller to trip is likely to cause others on the same power source to trip also.

If all spurious faults originate within controllers and interaction between controllers is carried on a common power supply circuit, the rate of false scrams is given by the expression of case A:

$$R_c = mnP_cF_s \tag{8}$$

where R_c is the rate of false scrams due to controller interaction, P_c is the probability that a spurious trip on one controller will be accompanied by a spurious trip on another, m is the number of controllers left which are eligible to trip the safety circuit after one has tripped, F_s is the rate of (natural) spurious faults, and n is the number of controllers in the circuit.

If all spurious faults originate on the common power supply and affect all controllers alike, the rate of false scrams is given by the expression of case B:

$$R_c = F_s$$

since every spurious fault trips every controller.

Cases A and B are evaluated in Table III. F_s was again assumed to be 12 spurious faults per year, and P_c was assumed to be 0.1. For a comparison, data from Table II, which ignored interaction, is included also.

The principal lesson to be learned from Table III is that none of the circuits can be considered reliable if interaction between controllers is present. Careful shielding and grounding and good basic circuit design are indicated.

Since a good controller will fail-safe or trip on loss of power, multiple sources of power are indicated for continuity of operation. Two sources are required for circuit 4, three sources for circuit 5, and four sources for circuit 6.

Maintenance of Safety Systems

Those charged with the responsiblity of maintaining a safety system can do many things to improve safety and reduce false scrams. For example, the period between testings τ appears in the denominator of the FM in Table I. Thus, it is important that the interval between tests be short if safety is to remain high. A regularly scheduled test once per shift can do much to insure that the equipment will respond properly when the need arises.

In general, routine maintenance should be staggered among the various controllers in a system in order to avoid a forced or coincident pattern of faults rather than the assumed random pattern. For example, the vacuum tubes in the electronic circuits should be checked and replaced on only one controller at a time. The probability of vacuum-tube failure is much higher during the first few hours after they are put in service.

A Proposed Design

The General Electric Company is designing the 180,000-kw Dresden nuclear power station for the Commonwealth Edison Company. Although the design is not complete, it has evolved to the point where it may be described in concept.

The safety circuit utilizes dual power

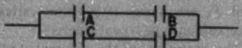


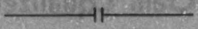
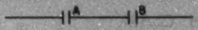
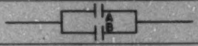
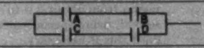
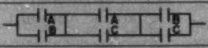
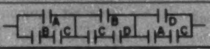
Fig. 4. Series-parallel circuit using four independent controllers

sources all the way to the rod-holding device. The schematic of Fig. 3 is a gross oversimplification of the portion of the circuit which holds the rod; nevertheless, it serves to explain the principle of the system.

One source of power is furnished to controllers *A*, *B*, and *C* and to one half of the safety circuit. Another independent source is furnished to controllers *D*, *E*, and *F*, and the other half of the safety circuit. Both safety circuits must open to scram the reactor. A trip on *A* or *B* or *C* in addition to a trip on *D* or *E* or *F* will scram the reactor. Failure of either source of power alone will not cause a scram. All three controllers in either group must suffer an operational fault before the reactor is unprotected. Two controllers, one in each group, must suffer a coincident spurious fault before the reactor is falsely scrammed. Interaction is expected to be kept at a minimum by the separate power supplies, isolated circuits, and careful grounding.

Of particular note is the fact that the system maintains its own spares in operation. If a controller should require repair, it is by-passed with the by-pass *BP* switch and removed from the circuit. The additional matrix of by-pass switch contacts are so interlocked that if more than one controller is removed from the circuit, that particular branch of the dual safety circuit is opened. This does not cause a scram; however, it does prevent the unsafe condition of operating with one controller that must trip before the reactor can scram. With one branch of the dual safety circuit interlocked open, the reactor operates with the single-series circuit. A high FM is maintained although the reactor is subject to a higher incidence of false scrams from spurious faults. It is expected that two controllers in one circuit by-passed at the same time for maintenance will be a rare occurrence.

TABLE III EFFECT OF CONTROLLER INTERACTION

CIRCUIT NUMBER	CIRCUIT ARRANGEMENT	CASE A R_c *		CASE D R_c *		R_s * *
1		F_s	12.0	F_s	12.0	12.0
2		2 F_s	24.0	F_s	12.0	24.0
3		2 P_c F_s	2.4	F_s	12.0	0.00546
4		8 P_c F_s	9.6	F_s	12.0	0.0279
5		6 P_c F_s	7.2	F_s	12.0	0.0164
6		12 P_c F_s	14.4	F_s	12.0	0.0323

* BASED ON 12 SPURIOUS FAULTS PER YEAR AND A P_c OF 0.1
* * COPIED FROM TABLE II FOR COMPARISON

Conclusions

The techniques of analysis outlined in this paper can never replace quality components and good design in safety systems, nor engineering judgment and experience in their use. As nuclear power plants join the ranks of power-producing utilities, increased emphasis will be placed on continuity of operation without sacrifice of safety. When the reliability of safety circuit amplifiers and controllers is advanced to a high state of perfection, but still is not good enough, these new techniques will find increased use in developing even better systems.

Safety systems designers may well take note that even though safety is the prime objective of a nuclear safety system, the record of false scrams is the basis on which the system will be most harshly judged.

Appendix

The series-parallel circuit using four independent controllers in the safety circuit will serve to illustrate the method of calculating the performance of typical circuits; see Fig. 4.

The false scrams per year can be determined as follows:

F_s = average rate at which controller *A* will develop spurious faults
$2F_s t_R$ = probability that either controller *C* or *D* will develop a spurious fault before controller *A* is reset

Thus, the rate of false scrams caused by controller *A* developing a spurious fault first is

$$R_A = 2F_s^2 t_R \tag{9}$$

Each controller will make a similar contribution to the rate of false scrams. Therefore the total rate is

$$R_s = 8F_s^2 t_R \tag{10}$$

The FM can be determined as follows:

$F_0 dt$ = probability that controller *A* will fault in the interval dt
$(\tau - t)$ = time remaining in the test period after controller *A* faults.
$F_0 t$ = probability that controller *B* will fault in the time interval before controller *A* faults

The contribution from controller *A* is

$$T_A = \int_0^\tau F_0^2 t(\tau - t)dt \tag{11}$$

Each controller can make a contribution equal to controller *A*. Thus the total contribution is

$$T = \int_0^\tau 4F_0^2 t(\tau - t)dt \tag{12}$$

which becomes

$$T = \frac{2F_0^2\tau^3}{3} \tag{13}$$

and the

$$\mathrm{FM} = \frac{3}{2F_0^2\tau^3} \tag{14}$$

A New Video Differential Phase-and-Gain Equalizer

J. H. CLARK
MEMBER AIEE

THIS PAPER describes a new technique for the improvement of color television circuits.

It is well known that the satisfactory transmission of National Television System Committee color television signals requires:

1. Flat gain-frequency response over the video band.
2. High signal-to-noise ratio.
3. Constant delay time versus frequency.
4. Good linearity, or constant gain or loss with respect to signal-level changes (differential gain).
5. Constant high-frequency delay time versus low-frequency signal level (differential phase shift).

The degree to which these requirements are met determines the relative merit of the color transmission circuit.

The same criteria apply to monochrome television transmission, except that monochrome is more tolerant with respect to differential gain and much more so for differential phase shift.

The terms differential gain and differential phase originated from work at the Bell Telephone Laboratories by S. Doba, R. C. Edson, H. P. Kelly, and others, and have become generally accepted as a convenient means of designating these two transmission characteristics. A good but brief discussion is found in a paper by Kelly,[1] which describes differential phase- and gain-measuring equipment.

Because of the different requirements, a good color circuit is always a good monochrome circuit, but a good monochrome circuit is not necessarily a good color circuit. The process of "colorizing" or improving a circuit by means of previous techniques consisted of two steps, as follows:

1. Design a good monochrome circuit with special attention to the flatness or amplitude-frequency response in the color carrier region from about 2.0 to 4.2 megacycles per second, using special equalization to accomplish this if necessary.
2. Reduce differential phase and gain effects by:
   a. Designing equipment to work well below the overload point.
   b. Using negative feedback.
   c. Selecting circuits that are inherently beneficial such as cathode-anode followers for high-level video, or special interstage networks for the intermediate-frequency portion of microwave relay equipment.
   d. Reducing the amplitude of the lower frequencies by the use of pre-emphasis and de-emphasis networks in sensitive portions of the circuit.
   e. Applying special delay equalizers to improve the characteristics of the intermediate-frequency amplifiers of frequency-modulated microwave relay equipment.

Although the design and adjustment of the delay equalizers mentioned in item e were very critical and exacting (especially for portable microwave equipment) by applying them, many satisfactory color circuits have been set up and used in the past several years. It was also possible, by the adjustment of these equalizers, to compensate at least partially for bad conditions elsewhere in the transmission path, but this compensation was difficult, time-consuming, and not always possible.

A more direct approach to the differential phase-and-gain problem was needed and this paper describes such a direct attack on the problem.

Fig. 1 shows the simplest form of the new equalizer, designed for 75-ohm input and output impedances and unbalanced operation. The attenuator shown is made variable in steps of 4 db (decibels) because this sometimes permits a better adjustment of the compensation curves, and also because it is usually desirable to use the minimum required loss, since this loss must be made up by gain external to this equalizer.

The center section of Fig. 1 contains the differential-gain adjustment which consists of diodes D_1 to D_4 and potentiometers P_1 and P_2. The type of diode is not critical. Type *1N56* has been used in all of the present commercial models. P_1 and P_2 are linear 10,000-ohm carbon-composition-type potentiometers.

The operation of the differential-gain section is very simple. Diode D_1 and the upper portion of P_1 control expansion of the waveform on positive peaks by shunting action on the series element of the attenuator. The characteristics of the correction curve are controlled by the input level, the setting of the attenuator, the slider of P_1, and the type of diode used.

The fitting of the correction curve to the distortion curve is observed by means of the equipment described by H. P. Kelly,[1] or by other similar equipment capable of measuring video differential phase and gain or of displaying the distortion curves on an oscilloscope.

Diode D_2 and the lower portion of P_1 provide compression on positive peaks by shunting the input. Usually the amount of correction required is not sufficient to cause any severe input impedance changes. Diodes D_3, D_4, and P_2 provide expansion or compression on negative peaks.

Usually the improvement is about 80% or more of the distortion. For example, a 1-db differential-gain impairment would be reduced to 0.2 db or less, and a 5-db impairment would be reduced to 1 db or less.

Greater improvement can be obtained, if required, by using additional diodes arranged with a small variable back bias and additional potentiometers, to secure a better fit for the compensation curve.

Another method of curve shaping is to use two diodes in series with a switch to short-circuit one diode when it is not required.

Note that this portion of the circuit is not frequency-sensitive, and it can be used for the correction of the nonlinearities of other circuits and devices in addition to the improvement of color television circuits.

Since the process is essentially one of waveform correction, it can be applied to improve nonlinear transducers such as camera tubes, Kinescopes and their associated circuits, audio equipment, including some types of microphone and loudspeaker distortion, as well as amplifier distortion, tape recording, wire photograph, and facsimile equipment. It is also of use in reducing intermodulation effects.

The correction may be applied either just ahead of or just after the device to be compensated, or at a distant point, provided that the delay-frequency curve is flat enough so that the fundamental frequencies and their significant harmonics

---

Paper 57-919, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, Pasco, Wash., August 28–30, 1957. Manuscript submitted May 20, 1957; made available for printing June 19, 1957.

J. H. CLARK is with the Pacific Telephone and Telegraph Company, Los Angeles, Calif.

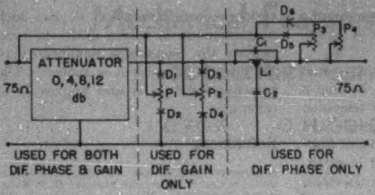


Fig. 1. Attenuator-type video differential phase-and-gain equalizer circuit

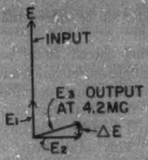


Fig. 2. Vector diagram of phase-equalizer voltages

do not substantially change their relative phase relationships. Delay equalizers can of course be employed to flatten the delay-frequency curve.

The right-hand portion of Fig. 1 contains the differential-phase corrector. C_1, C_2, and L_1 are designed as an all-pass network with a phase shift of approximately 90 degrees at 4.2 megacycles per second. The differential-phase adjustment is made by means of P_3 and diode D_5 for the positive signal peaks and P_4 with diode D_6 for the negative peaks. The conductance of the diodes in the forward direction is a function of the signal applied to each diode. P_3 and P_4 are in series with this variable conductance and also control, with the attenuator, the voltage level at each diode.

Referring to Fig. 2, vector E represents the input voltage E, and E_1 represents the voltage following the attenuator. At 4.2 megacycles per second, the phase-shift network rotates vector E_1 through an angle of 90 degrees to position E_2. At low frequencies there is little phase shift. A portion of E designated ΔE is combined with E_2, slightly increasing the voltage and shifting it to position E_3.

The magnitude of ΔE is a function of the signal level, the attenuation, and the setting of P_3 or P_4. ΔE therefore varies as the signal level changes and this results in a small controllable phase shift which is a function of signal level and polarity. This controllable phase shift is adjusted to compensate the differential phase shift of the over-all circuit.

In the example shown, the phase correction can be applied in only one direction. Fortunately, this is all that is usually required for the correction of video amplifiers or the intermediate-frequency amplifier of a frequency-modulated microwave radio-relay system of the type used for local television service. Other equalizers have been designed, using the same principle, to provide phase correction in both leading and lagging directions.

It was mentioned that the length of vector E_3 increased slightly as ΔE increased. Also, at low frequencies the effect of the differential-phase corrector is somewhat similar to that of the differential-gain corrector. Therefore it is desirable to adjust the differential phase first, since these effects will be included in the differential-gain correction, which is made last.

The amplitude-frequency irregularity compounded from the change in length of vector E_3 and possible losses in the phase network is of small magnitude and can be compensated by the variable equalizers, which are usually provided for final close adjustment of the over-all amplitude-frequency curve, or by a special single-section variable equalizer of the Bode or similar design.

The differential phase-and-gain equalizer shown in Fig. 1 is the simplest form which has been used for general application. Other circuits applying the same principles have been constructed, using an amplifier instead of (or with) an attenuator, which either have zero loss or provide usable gain in addition to the corrective function. It is evident that this type of corrector can be built into other devices instead of being applied as a separate piece of equipment.

Figs. 3 and 4 show commercial models of the attenuator type and zero-loss type respectively. The zero-loss type has been especially useful for cable video circuits.

Results

The results obtained from several months of testing, with more than a year of commercial operation, have been very good. It has been possible, by using this device, to dispense with all forms of intermediate-frequency delay equalizers for reasonably well designed microwave radio systems.

For example, a rather severe test was made on the monochrome radio-relay link set up for the Nevada atomic bomb television pickup, in May 1955. The color test was made on May 5, after the original 9-link circuit had been reduced to 7 links by by-passing two points because of the need for equipment elsewhere. The longest single link was 80 miles and the over-all length was about 320 miles using *RCA-X* type of equipment at all points.[3]

The circuit was tested during off-the-air time and no special adjustments were made. (Better results could have been had by special adjustments of the automatic-frequency control circuits on each link.)

The Kelly signal was sent at normal peak signal level from Nevada, and readings were made at Los Angeles. Approximately 50 degrees of differential phase

Fig. 3 (left). Attenuator-type equalizer

Fig. 4 (right). Zero-loss-type equalizer shift and 6 db of differential gain were measured.

Applying special 16-db pre-emphasis and de-emphasis networks, and using the new equalizer, just described, at Los Angeles, the differential gain measured 0.25 db and the differential phase shift 1.5 degrees. This adjustment would have provided a usable color television circuit.

Special tests through a television broadcast transmitter were also made with the co-operation of the National Broadcasting Company, and substantial improvement was obtained with a single equalizer on a built-up circuit containing three video cable loops, two microwave links, the studio equipment, the broadcast transmitter, and a Conrac off-the-air tuner. These tests showed irregular over-all differential phase-and-gain curves and indicated the desirability of using separate equalizers for different types of equipment.

In addition to the foregoing special tests, the new technique has become a part of the usual routine color television service of the Pacific Telephone and Telegraph Company in Southern California and has been employed in establishing many satisfactory color television services, both on video cable circuits and microwave radio systems.

Summary and Conclusions

A direct approach to the problem of constructing satisfactory color television circuits consists of using a new differential phase-and-gain equalizer. The new equalizer employs the nonlinear impedance characteristics of devices such as germanium or silicon diodes to compensate the undesirable nonlinearity of the circuits to which it is applied. The principle involved is also applicable to other uses.

References

1. Differential Phase and Gain Measurements in Color Television Systems, **H. P. Kelly.** *AIEE Transactions (Electrical Engineering)*, vol. 73, pt. I, Nov. 1954, pp. 565–69.

2. One Approach to a Video SHF Relay System, **R. H. Coe, F. F. McClatchie.** *Ibid.*, Sept. 1954, pp. 323–29.

3. Effect of Transmitter Characteristics on N.T.S.C. Color Television Signals, **G. L. Fredendall, W. C. Morrison.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., Jan. 1954, pp. 95–105.

4. The Measurement and Specification of Nonlinear Amplitude Response Characteristics in Television, **Stephen Doba, Jr.** *Ibid.*, Feb. 1957, pp. 161–65.

ERRATUM

"General Aspects of Hands-Free Telephony" by J. W. Emling, published in Communication and Electronics, May 1957, pages 201–05.

The author wishes to change the captions describing Figs. 6 and 7 on page 203 to read: "Oscillograms of voltage at set terminals from 50-millisecond pulses of 1,000-cycle tone . . . ." These pulses were referred to as 10-microsecond pulses in the published paper.

Mechanical Features of 561 Subscriber Carrier

J. E. ROSS
NONMEMBER AIEE

O. H. GOLDMAN
NONMEMBER AIEE

THE mechanical design of modern electronic equipment must meet two requirements: fabrication and assembly suited to modern production processes; and, equally important, ease of installation and maintenance. The Stromberg-Carlson *561*-type subscriber carrier was designed to meet both of these objectives. A companion paper[1] describes the electrical operation of this equipment. The discussion that follows deals with the mechanical problems only.

General Requirements

The requirements set forth in the following paragraphs governed the general design.

Since one end of the system was designed to serve the subscriber directly, all such equipment had to be located near or on the subscriber's premises. The outdoor operation imposed immediate requirements on the type of design to be followed and on the quality of the components.

The operational temperature range dictated by the limitations of transistors and electrolytic capacitors was established at −40 degrees Fahrenheit to +140 degrees Fahrenheit.

High ambient temperatures inside sealed cabinets exposed to the sun must be held within the upper extreme, since the germanium transistors and diodes used do not operate satisfactorily above this temperature. A higher temperature range could have been attained using silicon, but its use was deemed inadvisable for cost considerations.

Ruggedness and resistance to corrosion caused by humidity and salt were also considered as necessary conditions.

The handling of electronic gear during shipment and installation has always been a problem, and the equipment had to be tough enough to withstand rough treatment within practical limits.

Since in some instances vandalism may also be a problem, some type of protection against this had to be furnished on the pole-mounted unit.

Ease of installation embodying light weight, easy transporation, and a secure mounting of the pole box were also considered to be of prime importance to the buyer and consequently were to be accepted as real requirements in the design.

Possibility of servicing under all weather conditions is also a basic requirement. A pole box, which has as much protection from the elements as possible while open for service, was therefore considered mandatory. It was also felt that the lightning protection block should be externally accessible, so that replacement of carbon blocks could be effected without the necessity of opening the weatherproof box.

Designing the Package

The foregoing requirements illustrate in general the scope of the problem for the mechanical engineers working on this project. As work progressed, other related problems appeared. This paper discusses the general mechanical approach taken, which resulted in an end product meeting all the requirements and incorporating many new ideas never used before in carrier telephone field.

First and foremost was the problem of a rack-mounted housing for the equipment, since the more conventional methods were felt to be lacking in the degree of simplicity and flexibility required both for present and future production. After a thorough study of the available materials and processes, many ideas were discarded before selection was made.

Since the design required lightness, and die casting would not provide the desired flexibility of construction, sheet aluminum was chosen as the material providing evident advantages in weight, finish, and ease of fabrication as well as a clean-looking final product.

Ideas concerning construction and packaging were plentiful but though a prolonged search was made and several approaches considered, no satisfactory solution was apparent until printed wiring was considered. Through this medium a new line of thinking in regard to the packaging of the circuits became evident.

One might question the philosophy of using printed wiring in the packaging of this equipment instead of conventional point-to-point wiring. It is a well-known fact that the most costly, tedious, and time-consuming operation in the assembly of electronics or electric equipment is the hand assembly and hand soldering of the circuits. These operations are also the source of most of the errors which require costly inspection procedures.

Printing wiring gave rise to the grouping of circuit components on cards in such a manner that each functional circuit could be assigned a card, constituting a series in individual plug-in units as shown in Fig. 1. The over-all size was determined by the use of transistorized circuitry and human engineering. A rack height of $3^{1}/_{2}$ inches was decided upon since a smaller unit would have been difficult to service and a larger one wasteful of space and expensive.

The following arrangement was selected to fulfill the requirements and basic concepts stated in the foregoing.

The can or housing assembly which is used both at the central-office and subscriber ends is made up of two identical pieces each forming two sides of the housing, a 1-piece bottom, and identical brackets for rack mounting, spotwelded together to form the unit shown in Fig. 2.

Tolerance build-up, always the nightmare of the mechanical engineer, was held to a minimum by the utilization of counteracting tolerances. The problem of guiding the cards into their respective sockets was solved by the use of separate pieces of simple rectangular shape, sheared from strip stock and spotwelded in place, to form the guides. Through the use of these different width strips, and with a change in the indexing of the hole clusters for the sockets in the bottom of the assembly, the entire housing becomes flexible to suit the variations possible in equipment of this general type, thus reducing the need for new tooling and new design. Tolerance on the location of these strips is ±0.010 inch and on strip width is ±0.005 inch. Moderate tolerances, coupled with floating sockets for the cards, provide an adequate allowance for maximum warp in the cards while still ensuring positive plugging in without searching for the socket.

Paper 57-1017, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted June 7, 1957; made available for printing September 13, 1957.

J. E. Ross and O. H. Goldman are with the Stromberg-Carlson Company, Rochester, N. Y.

The authors wish to acknowledge the assistance provided by B. R. Stachiewicz, project engineer of Stromberg-Carlson Company, in the preparation of this paper.

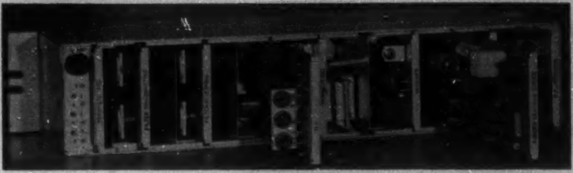


Fig. 1. Central-office terminal

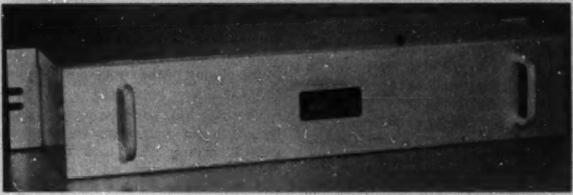


Fig. 2. Housing assembly

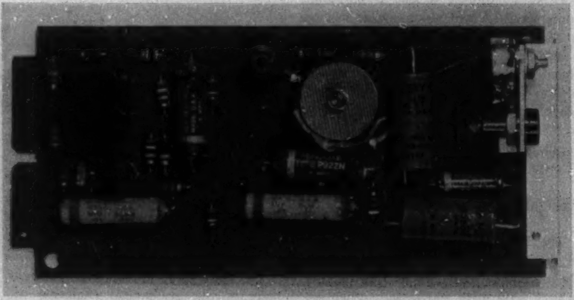


Fig. 3. Card assembly

Float for the sockets within reasonable limits was obtained by the use of headed pins with a body diameter smaller than the mounting holes in the socket. The pins are held in place by small retaining rings that are faster and easier to use than the conventional cotter key and washer. Further protection against binding is obtained by the use of stranded wire for the cable connected to the sockets. Each connection to each terminal is made with a small flexing loop to ensure that there be no bind caused by wiring.

All printed circuit cards and other units which may be added, moved, or replaced have been made plug-in. This provides simple and quick servicing by substitution. Plugs used on relays, oscillators, and similar components, with the exception of the cards, are standard and readily replaceable if broken. Sockets for all units are plainly marked and, in addition, those for the cards have polarizing keys to prevent mistakes during insertion. This is shown in Fig. 3.

Printed-Wiring Packaging

In the design of the printed-wiring cards, careful consideration was given to automation or mechanized assembly techniques. It might be well to point out at this time that Stromberg-Carlson Company has a complete mechanized production line for producing electronic assemblies of this type at the rate of 1,200 per hour; see Fig. 4. To utilize such production equipment properly, certain standards must be set up for the engineering department. These standards include such information as center-mounting distance of axial lead components and the inserting machines which can handle the various size components. Certain standard card sizes have been adopted, with tooling available. These standard card sizes were computed on the basis of the sheet sizes that can be purchsed economically and at the same time yield a minimum amount of waste.

The printed-wiring card material chosen, after careful consideration of all problems involved, is *XXXP* single-sided, 1-ounce copper. This material is a paper-base laminate that has high insulation resistance and low dielectric losses under severe humidity conditions, and presents very little manufacturing problems; consequently, it is economical to use. Its reliability is pointed out by the fact that it is being used in military equipment.

One of the biggest problems to be overcome was the warping of cards caused by the heat of dip soldering. This is an inherent characteristic of *XXXP* material. At one time, an epoxy-glass-base copper laminate was considered. However, this material did not have enough inherent advantages to offset the extra cost involved. Consequently, for reasons of economy, this approach was rejected. Two factors helped to reduce the warping problem encountered with *XXXP*. First, this material when purchased in a sheet form has a warp in a certain direction. By control of the printed-wiring card-manufacturing process, it is possible to produce the cards with the warp toward the short dimension. Second, by proper design of the plug-in portion of the chassis, compensation is made for the maximum amount of warping of the cards.

The printed-wiring board size (1/16 inch by 3.25 by 6.95 inches) is one of the standard sizes mentioned previously. It is rectangular in shape with the length being more than twice the width. Normally, this would be considered a poor design because a long narrow board will result in a great deal of warp along the length or long dimension. However, the warp problem was satisfactorily solved as previously stated. This, in fact, dictated the method used for plugging in the cards. To mount a connector plug on the end of the card would require a restricted guide made to close tolerances to accurately align with a receptacle.

Gold-plated contacts were used, etched in place directly on the board. The gold plating provided a good electric contact and protection against corrosion. The need for plating the contacts with nickel-rhodium was considered. However, the fact that the plug-in units were expected to function over a long period of time in a permanent position, the need for a connection providing a continuous wiping action was deemed unnecessary. Consequently, the nickel-rhodium, which is in effect more expensive than gold, was rejected.

Careful consideration was given to the shape of holes for mounting components on the board. The *561*-type subscriber carrier is the first commercially available high-grade communications equipment produced by an independent manufacturer incorporating the use of transistors and printed wiring. Although all theoretical information available to date indicates extremely long life of transistors, there has not been as yet a sufficiently long-term field experience in the use of these components. The expedient of plugging the transistors into sockets was therefore adopted, thus providing for easy field replacement. In the early design stages, the only holes in the board that required square corners were those necessary to mount transistor sockets. Fortunately, during the early design state, a new transistor socket became commercially available which would mount in round holes. As a result of this, all holes in the printed-wiring boards were made round. The only die work necessary is the notching of the corners at the plug-in end which can be accomplished by standard corner-notching dies and does not require special tooling. This was quite advantageous as it permitted pilot production runs without ordering costly dies requiring long procurement time.

Approximately 85% of the axial lead components, including some of the paper capacitors with eccentric leads, can be inserted by the mechanical equipment previously mentioned. As this constitutes a high percentage of all the components used, it is obvious that a high degree of mechanization could be achieved in the production of the printed-wiring assemblies.

The use of automatic production techniques also provided for a means of minimizing any damage to the semiconductors from heat during dip soldering, since the printed-wiring boards are dip-soldered with the solder temperature and the dipping cycle automatically controlled.

Some of the components that are specially made for this equipment are the potted transformers, of which there are three different types; see Fig. 3. The quantity of these components used does not warrant automatic insertion; consequently, the problem of proper orientation of these units was solved by offsetting one of the pins. This eliminated the possibility of human error during assembly. To increase the flexibility of the electronic circuits, it was necessary, in some instances, to provide a means for the use of factory-selected components. Wherever a selected component such as a resistor was specified, two terminal posts were mounted in the printed-wiring board. These terminal posts were placed as close to the edge of the board as possible, thereby allowing the operator to solder the leads without working among the other components. This, of course, minimized any possible damage that could be caused by the soldering iron.

It is of great importance in prefabricated circuitry to follow established electrical design principles, when applying printed-wiring packaging techniques, due to the small size usually associated with such assemblies. To eliminate problems introduced by stray capacitance or electrical noise, shielding of a printed conductor is accomplished by a printed shield consisting of conductor lines or areas

Fig. 4. Mechanized printed-wiring production

A—Card-exposing machine
B—Etch and neutralize machine
C—Automatic inserting line

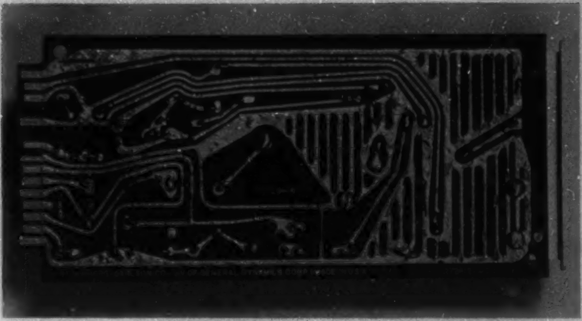


Fig. 5. Printed circuit card

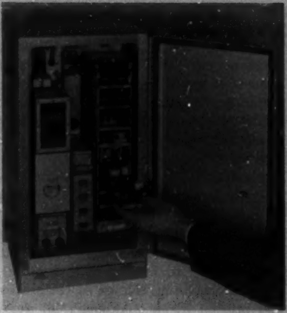


Fig. 6. Weatherproof housing and subscriber terminal

at ground potential on either side of the conductor to be shielded. A good illustration of this is shown in Fig. 5. By starting at the ground contacts, the ground shielding can be traced over the entire card. This illustration also points out how the large ground area pattern is broken up, to eliminate any unwanted coupling between portions of the circuit which can be caused by circulating currents in the ground area. It also reduces the concentration of heat in the dip-soldering operation, eliminating blisters.

To facilitate field service and maintenance, all printed-wiring card assemblies are designed with adjustable components, test jacks, and strapping posts (such as those on the voice-frequency receiver, fixed-frequency transmitter and high-frequency cards) located conveniently at the end of the card facing the cover. This permits field testing without necessitating the removal of the cards. To increase shelf life and alleviate field storage problems, all printed-wiring card assemblies are sprayed with an insulating material, providing a moisture and fungus proof surface with added insulation resistance.

The obstacles encountered in the design of the printed-wiring boards, such as a 3/16-inch blank space along each edge to clear the guides when the boards are inserted, blank 1/2-inch-square clearance in two opposite corners for pallet holes necessary for automatic insertion, and the extreme density of components on some of the boards, presented major design problems. However, complete design of all of the assemblies was accomplished with single-side copper laminate without the use of a single jumper wire.

Consequently, the printed-wiring medium of electronic packaging was adopted for this equipment with the following advantages realized:

1. Elimination of most of the hand assembly.
2. Elimination of most of the hand soldering.
3. Simplified assemblies.
4. Production time minimized.
5. Provision for a greater degree of miniaturization.
6. Inspection time was reduced and rejection rate held to a minimum.

The net result is a more uniform product with a high degree of reliability and considerable savings of time and space.

Central-Office Equipment

Total rack space required for a 5-channel system is 21 inches of standard 19-inch relay rack. It consists of 3 $^1/_2$ inches per channel and $3^1/_2$ inches for an extra panel equipped with a line coil and other filters. This results in a considerable saving of valuable space in exchanges. No servicing is required from the rear after the equipment is installed, which conserves additional floor area.

Card positions from left to right are: terminal card, filter cards (two), high-frequency card, transmitter card, receiver card, ringer card, compressor or voice-frequency transmitting card, expandor or voice-frequency receiving card. The oscillator, relay, and hybrid transformer are centrally located between the transmitter and receiver cards. The last two card positions mentioned accommodate the following optional combinations: compressor and expandor, or voice-frequency transmitting and voice-frequency receiving. No alteration in wiring is necessary when optional compandors are used.

Interchangeability of both central-office and subscriber ends is attained by making the maximum number of cards alike for all channels. Channel assignment may be changed by the insertion of the appropriate filters and an oscillator resulting in a completely versatile system which will allow the customer to use the same unit for any channel allocation at the central-office or at the subscriber terminal. This system allows stocking of a standard unit which can be manufactured on a highly mechanized production line with a resulting reduction in cost. The unit can then be stocked, withdrawn upon order for the insertion of the frequency-determining components, tested, and shipped with a minimum of effort.

Plug-in test cards are available which, when inserted in place of a card being tested, permit trouble shooting of the card while the system is in operation. In effect, the test card provides the means of extending the leads of the card being worked on while also supporting it in such a manner that the card will be clear of the housing.

Servicing by substitution can be accomplished by switching like cards from unit to unit or inserting a spare in the place of the "suspected" one. An extra set of cards may be kept on hand and a good circuit substituted in place of the faulty one while the faulty card is being serviced. This method allows the defective unit to be returned to the factory for service or replacement without loss of time and revenue to the telephone company. In some cases, if technical help of sufficient quality is available, the card may be serviced locally.

Pole-Mounted Equipment

It was necessary for outdoor mounting of subscriber terminals to provide a weather-sealed pole box of a construction that would remain reasonably simple, light, and require little effort to open for servicing. Lightness in weight is attained by the use of aluminum of heliarc-welded construction, to ensure against leaks in the body of the box.

The construction of an adequate and inexpensive door seal that could be bolted and would contain a gasketed flange proved to be a real problem. Again a search was made of types existing on the market and most of them were discarded as being too expensive or too poor for use. The problem was solved by the use of a piano-type hinge provided along the right-hand vertical edge of the door with two combination screw- and swing-type latches on the opposite edge for closing and applying pressure against the sealing element. The seal itself consists of closed cell neoprene mounted on the inside flat surface of the door and bearing against double edges on the box, providing a seal similar to the familiar labyrinth. Door edges overlap the box to provide a path to slow down wind-driven rain so that the seal is not exposed directly to the elements. The thin edges bearing against the seal give an added benefit in that little pressure is required to close the door to the point where the hold-down screws can take effect.

Protection against damage and the ability to dissipate heat are also important in this design and are obtained by the use of an external protective housing over and around the inner equipment housing. This "wrap-around," as it is called, is of 0.093-inch steel, plated and painted to provide durability, and by actual test this casing will stop a 0.22-caliber long-rifle bullet. The wrap-around is so constructed as to furnish an air space between itself and the inner housing, creating a chimney effect as well as deflecting the rays of the sun. Tests have shown the inside of the inner housing to be some 20 degrees Fahrenheit cooler with wrap-around in place.

The complete weatherproof housing consisting of the inner housing and wrap-around will be used for applications where extreme temperatures and/or durability are a factor, while the inner housing alone may be used in more temperate conditions.

The front of the outer housing is divided into two pieces, the upper overlapping the lower. A system of pins and slots is used to provide access for servicing. In actual service the upper half is raised to a horizontal position and latched to provide protection of equipment from the weather. It may also be slid back out of the way for easier access to the protection block which is mounted as an integral part of the inner housing. The lower half lets down to a level position to form a support or shelf for tools and instruments. When the inner housing door is opened, a 3-sided protection from the weather is provided while equipment is being worked on. The points herein described are illustrated in Fig. 6.

Lightning protection for the subscriber terminal is in the form of a built-in standard block manufactured by a nationally known concern. Access to this block is gained by opening the wrap-around, if one is used, and lifting a drip-proof cover at the top of the box. The inner housing need not be open during this operation. Protection of the 110-volt line is accomplished by the use of a pellet-type protector, which breaks down at around 175 volts, and is fastened in one of the knock-out holes in the main switchbox on the pole.

Conclusions

The mechanical design of the *561* subscriber carrier complements its electrical design. The use of transistors and printed wiring sets a precedent in the field of carrier-equipment engineering.

Unitized plug-in construction together with reliability and ruggedness make it unnecessary to assign trained personnel for maintenance or installation.

The authors feel that the design is practical and meets the requirements of high-grade rural service.

References

1. ENGINEERING FEATURES AND FIELD TRIAL PERFORMANCE OF A NEW SUBSCRIBER CARRIER SYSTEM, R. L. Layburn. *AIEE Transactions*, vol. 76, pt. I, 1957 (Jan. 1958 section), pp. 681–87.

2. A NEW 5-CHANNEL CARRIER SYSTEM FOR SUBSCRIBER USE EMPLOYING TRANSISTORS, B. R. Stachiewics, R. L. Layburn, D. F. Jamieson. *Ibid.*, Sept. 1957, pp. 385–90.

Engineering Features and Field Trial Performance of a New Subscriber Carrier System

R. L. LAYBURN
NONMEMBER AIEE

IN LATE WINTER of 1956, Stromberg-Carlson Company placed a new subscriber carrier system, known as *561* carrier, on field trial operation. This system is a result of approximately 15 months of intensive effort by many people in a number of specialized fields of engineering. System layout, circuitry, filters, packaging, printed circuitry, and components all required much basic engineering to achieve the results finally obtained in this system.

Prior to field trials, this system was given laboratory environmental tests over the temperature range of −40 to +60 C (degrees centigrade) and up to 95% relative humidity. Upon completion of all laboratory testing, a field trial was desired to demonstrate the effectiveness of the equipment under actual operating conditions. The results of this field trial, as well as a detailing of some engineering features, is the subject of discussion in this paper.

Engineering Features

561 carrier is a 5-channel rural open-wire double-sideband stackable system operating over the frequency range

---

Paper 57-1018, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted June 7, 1957; made available for printing August 13, 1957.

R. L. LAYBURN is with the Stromberg-Carlson Company, Division of General Dynamics Corporation, Rochester, N. Y.

The author wishes to acknowledge the assistance provided by B. R. Stachiewicz, project engineer, and D. F. Jamieson, circuit engineer, of Stromberg-Carlson Company in the preparation of this paper.

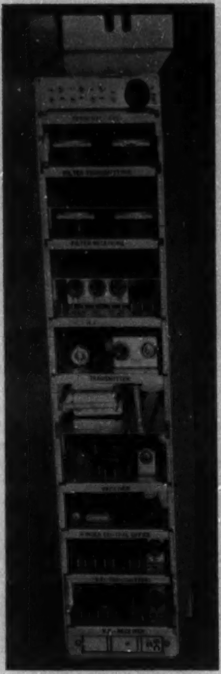


Fig. 1. Carrier unit

from 24.5 to 137.5 kc. All circuits, including transmitter, oscillator, receiver, ringing and dialing amplifiers, and regulated power supplies are completely transistorized. Each complete two-way channel utilizes 20 transistors.

The equipment is powered by −48-volt battery at the central office and by 117-volt a-c battery, 60 cps (cycles per second), at the subscriber terminal. Each central-office installation requires a maximum of 3 watts of idle power and 3.5 watts of operating power per channel. Each subscriber terminal requires 13 watts of idle power and 21 watts of operating power from the a-c line per channel.

Dialing and ringing are accomplished by pulsing the carrier. Ringing voltage, regenerated at the subscriber terminal, covers the harmonic- synchromonic- and decimonic-frequency ranges and can be applied to the subscriber drop on either a bridged or divided ringing basis. The equipment operates with all revertive call circuits that do not require simultaneous signaling on tip and ring. Provision is made for coded ringing and the system can handle as many subscribers per channel as the central-office equipment can code, on a maximum 1,200-ohm loop. Dialing pulses from 8 to 14 pulses per second are reproduced, with negligible distortion at the central office.

Fig. 2. Relay power panel

The transmitted carrier nominal output level is +5 dbm (decibels referred to one milliwatt) and the nominal receiver input level is −18 dbm. The equipment will operate within all specifications over the temperature range of −40 to +60 C and over line losses up to 35 db (decibels). With some degradation of performance at extreme temperatures, the equipment will operate up to line losses of 40 db. The receiver has a regulation range of 15 db over which the voice-frequency (v-f) output is constant within ±1 db. The equipment is designed to operate with a channel net loss of 4 db, between 2-wire drops.

In establishing the basic design of the system, particular emphasis was placed on ease of setup. The *561* carrier system can be added to the existing outside plant and central-office exchange facilities without any modification other than the addition of low-pass filters at existing drop points on the physical line to prevent loss to carrier signals. There

Fig. 3. Field trial equipped pole

is a single potentiometer in the carrier unit, which is used for adjusting the carrier level. In addition, there are three strap-in pads per carrier unit. Two of these are used to adjust v-f levels and are adjusted for nominal levels during manufacture. The third is adjusted in the field in accordance with line conditions. The v-f adjusting pads are mounted on two cards known respectively as the transmitting v-f card and the receiving v-f card. When required, compandors can be installed readily by replacing the transmitting and receiving v-f cards with the compressor and expandor cards.

The equipment was designed with a view toward minimizing and simplifying maintenance. This is achieved by arranging all circuits on printed-wiring plug-in cards. Also, all relays, terminal cards, line networks, and oscillators are plug-in type. All cards of each type, including transmitter and receiver, are interchangeable except for the transmitting and receiving filters. All plug-in cards are manufactured with gold-plated contacts to provide maximum contact reliability. The subject of mechanical design is covered in detail in another paper.[1]

During 6 months of field trial operation thus far, preceeded by almost 1 year of laboratory tests under maximum temperature extremes and exposure to maximum humidity conditions, a minimum amount of maintenance effort has been required to keep the equipment in proper operating condition. To date no transistor failures have been encountered. One power-supply diode failure was experienced during the height of the thunderstorm season because of failure of the 117-volt power-line protection to

Fig. 4. Field trial central-office installation

perform adequately. Power-line protection is generally recognized as a difficult problem. Because of lack of recurrence of this failure, it may be classified as a chance occurrence and a slight unavoidable risk.

The channel transmission and signaling equipment required at the central office is packaged in an aluminum housing designated as the carrier unit, occupying $3^{1}/_{2}$ inches of standard rack space. One such unit is required at each terminal. This is shown in Fig. 1. At the subscriber terminal, another $3^{1}/_{2}$-inch panel is needed to perform switching, ringing regeneration, and power-supply functions. This is shown in Fig. 2. At the subscriber terminal, the carrier unit and the ringing-relay power panel are assembled in a pole-mounted box together with an auxiliary panel which mounts plug-in transformers and filters, and a-c power receptacles. Connections to the carrier line and the carrier-derived voice circuit are made at the top of the pole box. This is shown in Fig. 3. Lightning protection is provided within the terminal block. A protective housing made of steel, as shown in Fig. 3, will be provided with the pole box, where required, to protect against excessive temperature rise from sun rays and against physical damage to the pole box.

At the central office, the carrier units are mounted on standard 19-inch racks. Fig. 4 shows a complete 5-channel central-office installation. One additional panel, shown mounted at the top, is required to accommodate the line transformer, the low-pass filter required for isolation of existing voice circuit, and filtering for the central-office battery. The jack field shown below the carrier units is not customarily required, but was used during the field trial to provide line circuits, line terminations, and test circuits.

Fig. 5. Carrier terminal and stand-by power pole boxes on field trial

Where required for service reliability, stand by power is provided in a pole box of a size approximately equal to the carrier pole box. This is shown in Fig. 5, mounted on a pair of crossarms adjacent to a carrier pole box. The power supply consists of 23 lead-calcium cells floating across a charger. The floating voltage is −53 volts over the temperature range of +5 to +60 C. Below +5 C, the stand-by power box is heated using thermostatic control to assure maximum battery discharge efficiency. The box is insulated to limit the power drain required to keep the box warm and to hold in heat upon a-c power failure. The batteries permit normal usage of the carrier equipment for at least 40 hours after a-c power failure for a single-channel installation at temperatures above +5 C and for at least 8 hours at −40 C. Should the power failure continue beyond this period, the cells will not freeze because of the high specific gravity of the electrolyte.

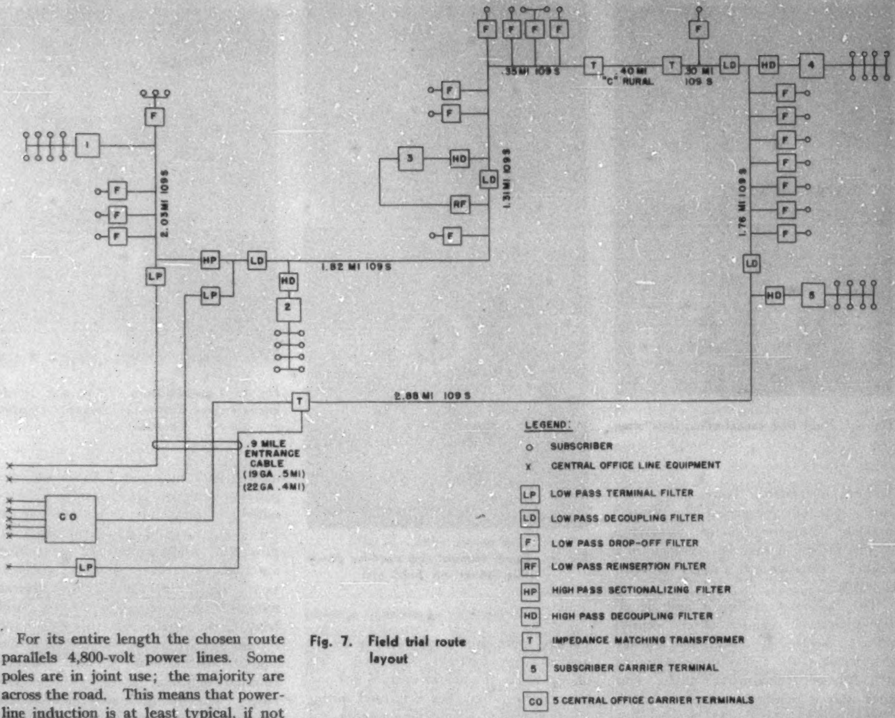
Fig. 6. Low-pass drop-off filter and impedance-matching transformer on upper crossarm on field trial

V-f low-pass drop-off filters, sectionalizing 4-kc high-pass filters, and impedance-matching transformers (designated *F*, *HP*, and *T* respectively in Fig. 7) potted in hard rubber casings are provided. These can readily be mounted on crossarms or other surfaces, as necessary. Fig. 6 shows two such networks mounted on a crossarm for the field trial. A soft rubber cover slips over the top and sides of the network for protection against weather and physical damage. Internal lightning protectors are provided. The protectors are latticed across the filter terminals in a manner which bypasses surges around the network. Ground wires are not required in this arrangement.

Field Trial Layout

In selecting a field trial site, particular attention was directed toward securing a long line which would be typical of rural circuits. The line finally selected fits these requirements very well, although the setup is unique in the manner in which it was accomplished. The route selected includes 0.9 mile of entrance cable. This is considered a necessary part of a test route to permit evaluation of higher carrier-frequency operation and to determine the effect of central-office switching transients on carrier channel noise. In actual use most systems will have to operate over some length of entrance cable.

On the open-wire section, typical construction (including for the most part 109 iron wire with 12-inch spacing, *R-1* transposed) of sufficient length was desired and obtained.



.35 MI 109 S
.40 MI "C" RURAL
.30 MI 109 S
2.03 MI 109 S
1.31 MI 109 S
1.82 MI 109 S
1.76 MI 109 S
2.88 MI 109 S
.9 MILE ENTRANCE CABLE (19 GA .5 MI) (22 GA .4 MI)
1
2
3
4
5
CO
LEGEND:
○ SUBSCRIBER
x CENTRAL OFFICE LINE EQUIPMENT
LP LOW PASS TERMINAL FILTER
LD LOW PASS DECOUPLING FILTER
F LOW PASS DROP-OFF FILTER
RF LOW PASS REINSERTION FILTER
HP HIGH PASS SECTIONALIZING FILTER
HD HIGH PASS DECOUPLING FILTER
T IMPEDANCE MATCHING TRANSFORMER
5 SUBSCRIBER CARRIER TERMINAL
CO 5 CENTRAL OFFICE CARRIER TERMINALS

Fig. 7. Field trial route layout

For its entire length the chosen route parallels 4,800-volt power lines. Some poles are in joint use; the majority are across the road. This means that power-line induction is at least typical, if not more severe than would be encountered in the average case.

The increase in length of the open-wire circuit was achieved by combining three separate voice circuits into one carrier line. This was accomplished, as shown in Fig. 7, by interconnecting pairs normally carrying voice through sectionalizing 4-kc high-pass filters (*HP*) and isolating existing v-f equipment with v-f low-pass filters (*LP*). Tests of the completed circuit indicated that the total carrier path transmission and impedance characteristics were not affected in any material way by this manner of achieving a long carrier line.

The three v-f segments of the line are quite typical of a subscriber carrier installation wherein physical service is maintained on the initial portion of the route and voice frequencies derived from two carrier channels are reinserted on the carrier line for transmission to subscribers' premises. Note that a total of 18 low-pass filters are used across the line for dropping off subscribers. Of these filters eight were used for actually serving subscribers on a reinserted basis and the remainder were subscribers served over the physical circuits.

The route also afforded an opportunity to try out the use of *C* rural distribution wire for a short segment of the carrier path. *C* rural distribution wire is recommended by manufacturers for short carrier lengths up to 2,000 feet. This opportunity presented itself when new construction was required to connect two v-f circuits separated by about 0.4 mile.

Fig. 8 gives the carrier-frequency assignments and distance from central office to each subscriber terminal. The transmission versus frequency characteristic of the carrier line between the central office and each subscriber terminal is shown in Fig. 9. In setting up the channels, it was intended to operate them under line conditions which would provide maximum allowable line loss under wet-weather conditions. The arrangement was purposely made to simulate a rather severe application to permit a more precise evaluation of system performance, under worst line conditions, than was possible in laboratory tests.

Note that impedance-matching transformers (designated *T* in Fig. 7) were used at the open-wire junction and at the *C*-wire junctions. High-pass and low-pass filter combinations (designated *HD*, *LD* in Fig. 7) were used at each subscriber terminal. These filters, consisting of half-section networks adjusted to an appropriate cut-off point for each pair of channel carrier frequencies, are simple, low-cost plug-in units. These filters are used when it is desired to terminate the carrier line for each carrier frequency to avoid bridging loss. Another advantage to be gained in this arrangement is a reduction in crosstalk and noise due to the effective decoupling accomplished between succeeding subscriber terminals.

Bridging of each subscriber terminal across a carrier-terminated line was also tried. It is recognized that in some

CHAN	TRSG FROM CO	TRSG FROM SUB	DISTANCE TO EACH SUB TERM MILES *
1	27KC	39KC	10.85
2	51	63	8.82
3	75	87	6.19
4	99	111	4.64
5	123	135	3.25

* PLUS 0.9 ENTRANCE CABLE

Fig. 8. Frequency allocation and distance to each subscriber terminal in field trial

cases this connection would be desirable from an economy standpoint. The results were as would be expected, with the channel-decoupling method demonstrating the advantages listed in the foregoing. Approximately 3½-db bridging loss in all channels was experienced, whereas with decoupling this loss was experienced only in the most remote terminal which requires transmission through four low-pass decoupling networks.

Primary lightning protection with 0.006-inch gaps was placed on the second and fifth pole away, on each side of all subscriber terminals. In addition, 175-volt protection was placed on the 117-volt a-c incoming power leads. This protection, in addition to that provided in the subscriber terminal block, is considered adequate to protect the carrier terminal against all but direct hits on the equipment. All protectors used in the system are of the self-restoring type.

Because engineering tests would interfere with normal service, subscribers were not connected until completion of these first tests. The desirability of being able to make additional engineering tests at any future time was recognized; therefore, subscriber connections were made in such a manner that they could readily be restored to service over the physical lines. For this purpose a cutover switch was utilized in the central office, and strapping arrangements were made in fuse boxes at subscriber terminals. A terminal-per-station central office was selected for field trial so that these changes could be made at the main distributing frame without changing customers' numbers. One channel provided service to subscribers by distribution of voice frequency over the carrier line. The other channels served subscribers on other pairs in the same route.

Field Trial Results

The regulation characteristic for a typical carrier channel is charted in Fig. 10. Note that the equipment regulates over an input range of +5 to −10 db from a nominal receiver input of −18 dbm (corresponding to line and fixed losses of 23 db) with a constancy of v-f output of better than ±1 db. This range of regulation has been accomplished with a new transistor circuit utilizing negative resistance elements connected in such a way as to provide backward-acting regulating action.

The detector transistor used in this circuit presented many problems before a suitable solution was obtained. The requirements of current and voltage are such that over 300 milliwatts of collector dissipation result, thus requiring that a power-transistor type be utilized. This presented a problem in obtaining a power transistor with adequate high-frequency current amplification, high maximum collector voltage, and low collector cut-off current. A maximum cut-off current of 150 microamperes is permitted to limit bias on the negative resistance elements connected in the collector circuit. Although these requirements were almost prohibitive, a satisfactory solution was achieved through close co-operation with a transistor manufacturer. The common-emitter configuration has proved very satisfactory because the current gain achieved enables excellent regulation of receiver output over the operable range of the receiver. At the same time, however, the emitter resistance increases greatly with very small signals and as a result approximately 20-db loss of detection efficiency to small unwanted signals is realized. This reduces crosstalk and noise at the output of the channel as indicated in Fig. 11. The improvement obtained varies from channel to channel and can only be approximated as a general improvement figure.

Equipment noise in terminals operated back to back without compandors does not exceed 26 dba (decibels adjusted); crosstalk under the same conditions is held below 29 dba with 0-dbm disturbing level. Both noise and crosstalk figures are referred to zero transmission level. The use of compandors is optional, therefore, and would be recommended only where required by line conditions. Channel equipment is normally equipped with v-f transmitting and receiving cards containing adjusting pads, test points, and low-pass filters. In the event the compandor option is selected, these cards are replaced respectively by the compressor and expandor.

The idle noise which can be expected from the five channels as set up in the field trial without compandors is computed in Fig. 11 and is compared against average measured values. The calculations are based on meager noise data that the author was able to gather. These data are based on measurements of typical rural lines in various parts of the United States. The values given are approximate in nature primarily due to the relatively short lengths of line that were measured. However, the results are sufficient to indicate that the system noise output for given line conditions is within expectations.

Compandors have been installed in the field trial and are yielding the anticipated results in suppressing the effects of high line noise. Fig. 11 provides a comparison of noise experienced with and without compandors.

Radio transmissions were picked up in

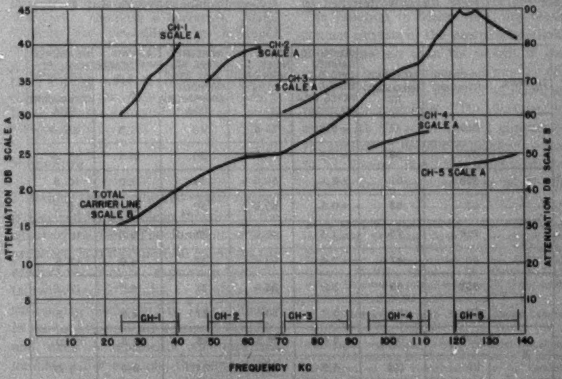


Fig. 9. Field trial carrier line loss versus frequency

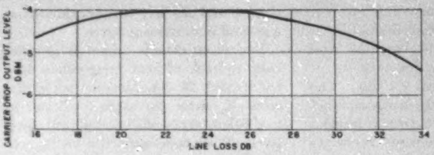


Fig. 10. Receiver regulation

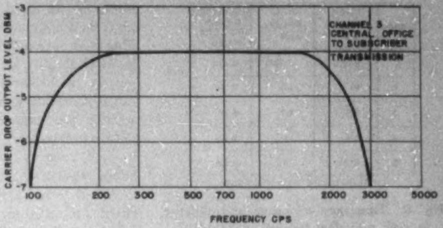


Fig. 12 (right). Frequency-response drop to drop for 4-db net loss circuit

some channels during the field trial. This interference proved troublesome without compandors, particularly in one channel in which the interference fell inband. In some circumstances such interference is sufficient to preclude normal operation of carrier channels unless extraordinary measures are taken. Unbalance of the carrier line and equipment was investigated and was found to be satisfactory. Known methods of noise suppression were reviewed and it was concluded that compandors would provide the best solution for reducing this radio interference. Compandors proved to be very satisfactory. In the worst case the noise level attributable to radio transmission was reduced from 40 dba to less than 20 dba.

In a subscriber carrier system the physical circuit must be given consideration as a potential noise source. Ringing from either the central office or regenerated and reinserted ringing of the carrier terminal present somewhat the same problem. Generally the latter has a somewhat better waveform and presents a lesser problem. The low-pass terminal filters and drop-off (way station) filters (see Fig. 6) were specially designed to reduce this interference. Dialing on the same line over which carrier is being transmitted also causes interference in the carrier terminals. This effect is caused by the opening and closing of dial contacts, the former causing sparking and the latter resulting in a momentary short-circuit across the line. The Stromberg-Carlson *F1543W*-type telephone has been designed to operate with the system, and to provide minimum dialing interference within the objectives set up for the system.

Fig. 12 shows a typical frequency response from drop to drop for one channel, including transmission through all line equipment. The receiver output amplifier is a single-transistor common-emitter amplifier using adequate feedback to provide a flat response and has an output power of +6 dbm over the desired frequency range and an output impedance of 600 ohms. A low-pass filter limits carrier leak. The response shown is within the −3- to +1-db requirement of the equipment, from 250 to 2,500 cps. Distortion for all channels at 250 cps is an average of 2.5%.

Ringing is regenerated at the subscriber terminal through amplification of detected carrier pulses which operate a mercury relay. Divided ringing is handled by turning on a tone generator in the central-office carrier unit when ringing is on the tip side of the line. This operates a relay at the subscriber terminal, which grounds the ring side and connects the tip side to the ringing generator.

CHANNEL	RECEIVING TERMINAL	RECEIVING FREQUENCY KC	COMPUTED CARRIER LINE 5 % STATIC NOISE DBA	COMPUTED NOISE AT CARR DROP DBA	MEASURED NOISE AT CARR DROP (WITHOUT COMPANDORS) DBA	APPROX. IMPROVEMENT IN S/N DB	MEASURED NOISE AT CARR DROP (WITH COMPANDORS) DBA
1	SUB	27	+8.6	51.5	40	11.5	20.0
	CO	39	+4.2	54.5	30	24.5	18.5
2	SUB	51	+2.1	50.0	38	12.0	16.5
	CO	63	+0.3	50.0	30	20.0	18.0
3	SUB	75	−1.3	42.0	33	9.0	18.0
	CO	87	−2.4	43.0	32	11.0	17.5
4	SUB	99	−3.2	35.0	31	4.0	17.0
	CO	111	−3.7	35.5	27	8.5	15.5
5	SUB	123	−4.4	31.5	25	6.5	15.0
	CO	135	−4.9	32.0	24	8.0	15.0

Fig. 11. Field trial noise table (FIA rating)

When the mercury relay is operated on a 40% to 60% duty cycle, the harmonic content of the ringing output will be less than 30% which is adequate to ring only the desired telephone on a party line. During field trials before subscribers were connected, the operation of the ringing relay was measured with a per-cent make meter to assess the quality of regenerated ringing. This proved to be between 44% to 50% in all channels up to 40 db of line loss, for ringing frequencies from 16 $^2/_3$ to 66 $^2/_3$ cps.

Dialing distortion was measured during field trial tests by comparing the operating distortion of a battery-feed pulsing relay when operated directly from pulsing contacts of a given per-cent make and when operated with the carrier terminals interspersed between the two. In the latter case the line-circuit relay is operated by closing its circuit through the hybrid coil of the carrier equipment and through contacts of the dialing relay. The dialing amplifier is a conventional common-emitter stage which picks up carrier signals from the receiver before detection. The dialing pulses are passed through a voltage doubler and then to a dialing-relay amplifier. Dialing distortion was measured with line losses up to 40 db. In no instance did distortion exceed 4%.

Conclusions

The field trial demonstrates that the *561* subscriber carrier system provides an answer to the need for a high-quality easily maintained carrier system to be used on rural subscriber lines. This carrier operates between central office and subscriber locations and performs all supervisory, dialing, and ringing functions satisfactorily. Transmission quality is high, and noise results are within specifications. In several months of field operation, no component failures of any kind point to a high degree of reliability. Although a full thunderstorm season has not yet passed, several severe thunderstorms have occurred with no transistor or other component failures.

Completely satisfactory subscriber service should result, leading to increased revenues per dollar of outside plant for telephone operating companies.

References

1. A New 5-Channel Carrier System for Subscriber Use Employing Transistors, B. R. Stachiewicz, R. L. Layburn, D. F. Jamieson. *AIEE Transactions*, vol. 76, pt. I, Sept. 1957, pp. 385–90.

2. Mechanical Features of 561 Subscriber Carrier, J. E. Ross, O. H. Goldman. *Ibid.*, vol. 76, pt. I, 1957 (Jan. 1958 section), pp. 677–81.

A Numerical-Graphical Method for Synthesizing Switching Circuits

A. H. SCHEINMAN
NONMEMBER AIEE

ALTHOUGH methods for synthesizing switching circuits have been developed, the search for easily applied techniques is still receiving considerable attention. Recently, for example, a great deal of work has been done on the development of matrix methods,[1] which lead to efficient nonseries-parallel circuit arrangements. Matrix methods, however, are still in the formative stage and have not as yet proved to be practical circuit-design tools. The direct application of Boolean algebra,[2] on the other hand, yields circuits of the series-parallel type, which generally are not very efficient. In addition, a designer has a large number of alternative procedures available to him in the application of Boolean algebra, and can not usually ascertain beforehand the procedure which will result in the best circuit arrangement.

Graphical techniques seem to offer decided possibilities in the search for efficient circuit-design methods. Indeed, considerable effort has been expended in this direction.[3,4] The best of the graphical techniques at present seems to be Karnaugh's map method. The map method has a major limitation, however, in that it rapidly exceeds the limits of practicability as the number of variables involved increases.

In actual practice it appears that most circuit designers simplify switching circuits by inspection based on experience. Perhaps one of the reasons for this is that present methods often become very difficult to apply to practical problems.

The method described in this paper consists of the visual comparison of a series of decimal numbers which represent the switching function to be simplified. Contacts of the switching variables involved are systematically extracted by the use of simple, easily applied rules. This method also has the following unique advantages:

1. It is not necessary to recognize that a function is symmetrical in order to realize the minimal circuit, since symmetrical circuits evolve automatically.
2. The method does not break down in the development of circuits having a large number of variables.

Since this method relies on a graphical procedure, it is explained largely by the use of illustrative examples.

In the switching algebra that appears in this paper, the following conventions are used:

0 = open circuit.
1 = closed circuit.
Addition (+) = parallel connection = or.
Multiplication = series connection = and.
An unprimed variable is used to denote a make contact (a contact of a relay which is closed when the relay is operated). Make contacts are symbolized by a small cross (×).
A primed variable is used to denote a break contact (a contact of a relay which is closed when the relay is not operated). Break contacts are symbolized by a small vertical line (|).

Procedure

Assume that the desired circuit output is specified by a Boolean function of the circuit variables. The function is then expanded into its canonical form and the product terms of the expression thus obtained placed in a table of combinations such as that shown in Fig. 1(A). Each 0 in the table represents a primed variable, and each 1 represents an unprimed variable. For example, the term $A'B'C$ corresponds to the entry 001.

Each row of the table has the appearance of a binary number and is therefore assumed to be a binary number. The decimal equivalent of each row is then formed. Each product term of the function is thus represented by a decimal equivalent. Note that each circuit variable now has a fixed decimal weight depending upon the order of the variables in the table. For example, in Fig. 1, A has a weight of 2^2 or 4, B has a weight of 2^1 or 2, and C has a weight of 2^0 or 1.

The Boolean function in its expanded form may therefore be expressed as a sum of decimal numbers. Thus, in Fig. 1, the function of the input variables A, B, C may be expressed as $F(A, B, C) = \Sigma 1, 2, 5, 6$. Henceforth, the summation sign is omitted.

The next step in the development of the circuit is to extract contacts of the input variables starting with the most heavily weighted variable. The function specified in Fig. 1 is completely developed as an illustration.

A make and a break contact of the most heavily weighted variable, A, is removed. In series with each contact of A, a new function is formed and designated as F_1 and F_2. F_1 and F_2 are functions of the variables B and C only, since the variable A has already been extracted. F_1 and F_2 are obtained in the following manner. The function in series with the break contact consists of the sum of all decimal equivalents less than the weight of the variable being extracted. Thus $F_1(B, C) = (1, 2)$ since the weight of A is four as previously explained. The function in series with the make contact consists of the sum of all decimal equivalents equal to or greater than the weight of the variable being removed, from which the

---

Paper 57-1019, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted May 22, 1957; made available for printing August 12, 1957.

A. H. Scheinman is with Bell Telephone Laboratories, Inc., New York, N. Y.

This paper represents, in part, material submitted to the Polytechnic Institute of Brooklyn in partial fulfillment of the requirements for a master's degree in Electrical Engineering. The author wishes to thank Prof. E. J. Smith of the Polytechnic Institute for his help and encouragement.

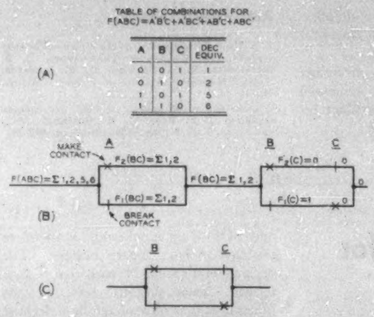
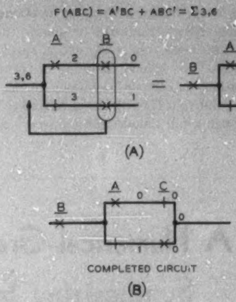


TABLE OF COMBINATIONS FOR
F(ABC)=A'B'C+A'BC'+AB'C+ABC'
(A)
MAKE CONTACT
A
$F_2(BC)=\Sigma 1,2$
F(ABC)=Σ 1,2,5,6
F(BC)=Σ 1,2
$F_1(BC)=\Sigma 1,2$
BREAK CONTACT
(B)
B
C
$F_2(C)=0$
0
0
$F_1(C)=1$
0
B
C
(C)
COMPLETED CIRCUIT

A	B	C	DEC EQUIV.
0	0	1	1
0	1	0	2
1	0	1	5
1	1	0	6

Fig. 1 (left). Illustration of rule 1

Fig. 3 (right). Illustration of rule 3


F(ABC) = A'BC + ABC' = Σ 3,6
A
B
2
0
3,6
3
1
=
B
A
0
1
(A)
A
C
0
0
B
0
1
0
COMPLETED CIRCUIT
(B)

weight of the variable is subtracted. Thus, subtract 4 from 5 and 6 and therefore $F_2(B, C)=(1, 2)$. Since $F_1=F_2$, they may be connected and thus combined into one function. This example was chosen to illustrate rule one, which is:

Rule 1. Equal functions of the same variables may be combined into a single function.

The remainder of the development of this illustration is self-evident. In this example the make and break of relay A are connected and therefore constitute a short-circuit as shown by the theorem $X+X'=1$.[2]

Note that the junction of a make and a break contact of a relay forms a transfer contact.

In the examples that follow, only the decimal equivalents are shown. It is to be understood, for example, that the decimal equivalents specifying the input function are functions of all of the circuit variables and the decimal equivalents resulting from removal of a variable are functions of all succeeding variables.

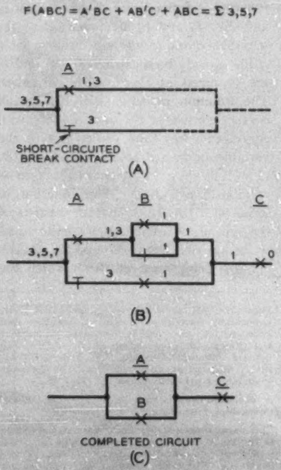

F(ABC) = A'BC + AB'C + ABC = Σ 3,5,7
A
1,3
3,5,7
3
SHORT-CIRCUITED BREAK CONTACT
(A)
A
B
C
1,3
1
1
1
3,5,7
1
0
3
1
(B)
A
B
C
COMPLETED CIRCUIT
(C)

Fig. 2. Illustration of rule 2

The next example illustrates rule 2. Assume an input function of (3, 5, 7) as shown in Fig. 2. Removal of a transfer of relay A results in the functions (3) and (1, 3). The break contact of A can be short-circuited as may be seen by performing the Boolean manipulation for the circuit:

$$A(Y+Z)+(A'Y)=(AZ)+Y$$

where Y represents the term 1 and Z represents the term 3. Thus the second rule may be expressed as follows:

Rule 2. If one of the two functions emanating from a transfer contact is a subset of the other, the element of the transfer in series with the subset may be short-circuited. The decimal numbers of the other function corresponding to those of the subset become "don't-care" terms.

Manipulation of don't-care terms is treated subsequently.

If a contact is to be short-circuited, a small line is placed over it to indicate this. The third rule is:

Rule 3. If in the development of the circuit, identical contact configurations appear in series with the make and break elements of the transfer contact being removed, they may be shifted to the left of the transfer contact and combined.

Figs. 3 and 4 illustrate rule 3. In Fig. 3 the identical configurations are single make contacts of relay B, while Fig. 4 shows a slightly more complicated case where the identical configurations consist of transfer contacts of relays B and C.

Rule 3 is permissible because of the commutative and distributive laws of the Boolean algebra. In addition, disjunctivity is maintained because of the presence of the transfer contact.

Don't-Care Terms

A don't-care term is a product term for which a circuit closure is arbitrary. In practical problems these may exist because the particular combination of variables of the product term involved can not occur. This method handles don't care or d terms as follows:

All decimal equivalents of d terms have a bar placed over them to differentiate them from other terms. Numbers which have bars over them (d numbers) are treated in the same manner as any other numbers. If, when a transfer contact of a relay is removed, and all of the decimal numbers in series with one of its elements are d numbers, then that element, and all of the d numbers in series with it, may be removed at this time, since no further simplification will result by their retention. The exception to this general rule is the case where by retaining all or some of the d numbers the function may be made equal to that of the other transfer element, thus eliminating that transfer contact. Note that in combining two functions by rule 1, if corresponding terms of each function are d numbers, the corresponding term of the combined function is a d number. But, if corresponding terms are d numbers in one

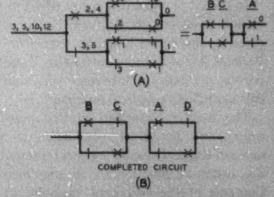

F(ABCD) = A'B'CD + A'BC'D + AB'CD' + ABC'D' = Σ 3, 5, 10, 12
A
B
C
2, 4
0
0
0
2
0
3, 5, 10, 12
3, 5
1
1
1
3
1
=
B C
A
0
1
(A)
B
C
A
D
COMPLETED CIRCUIT
(B)

Fig. 4. Illustration of rule 3

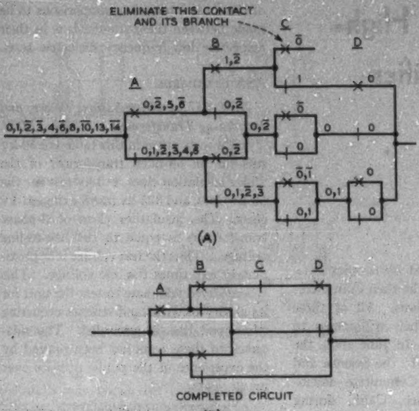

ELIMINATE THIS CONTACT AND ITS BRANCH
A
B
C
D
$0,1,\bar{2},\bar{3},4,\bar{6},8,\overline{10},13,\overline{14}$
$0,\bar{2},5,\bar{6}$
$0,1,\bar{2},\bar{3},4,\bar{6}$
$1,\bar{2}$
$0,\bar{2}$
$\bar{0}$
1
0
$\bar{0},\bar{1}$
$0,1,\bar{2},\bar{3}$
0,1
(A)
COMPLETED CIRCUIT
(B)

Fig. 5. Manipulations with don't-care terms

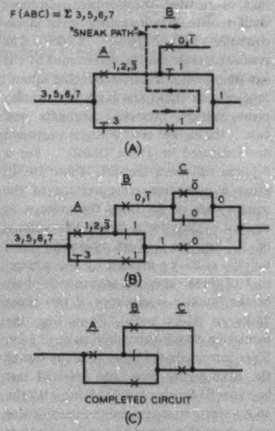

F (ABC) = Σ 3, 5, 6, 7
"SNEAK PATH"
A
B
C
$1,2,\bar{3}$
$0,\bar{1}$
1
3
3, 5, 6, 7
(A)
$\bar{0}$
0
(B)
COMPLETED CIRCUIT
(C)

Fig. 6. Illustration of sneak path

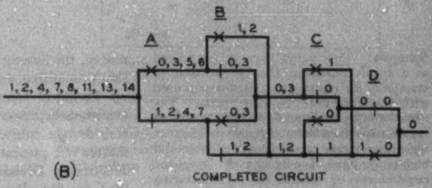
function and not in the other, then that term is not a *d* number in the combined function.

Fig. 5 illustrates these operations.

TABLE OF COMBINATIONS

A	B	C	D	DEC. EQUIV.
0	0	0	1	1
0	0	1	0	2
0	1	0	0	4
0	1	1	1	7
1	0	0	0	8
1	0	1	1	11
1	1	0	1	13
1	1	1	0	14

(A)


A
B
C
D
1, 2, 4, 7, 8, 11, 13, 14
0, 3, 5, 6
1, 2, 4, 7
1, 2
0, 3
0
1
COMPLETED CIRCUIT

(B)

Fig. 7. One or three out of four symmetric

Extraneous Circuit Closure

It is possible to introduce an extraneous circuit closure or "sneak path" by application of rule 2 when followed by application of rule 1 in one particular case. This is illustrated in Fig. 6. Therefore in applying the rules, rule 1 should be applied first. Rule 2 should be applied after rule 1, and if a sneak path results, rule 2 should not be applied.

Symmetric Circuit

As a final example, a moderately difficult symmetric is developed to illustrate the ease with which the method handles this type of circuit. Consider a circuit of four relays, for which a circuit closure is desired when any one or three of the relays are operated. The table of combinations and the circuit development can readily be followed in Fig. 7.

Conclusions

This method enables a circuit designer to rapidly develop economical circuit arrangements by the systematic application of three simple rules to a numerical representation of a Boolean function. In addition, by changing the order of the variables in the table of combinations (which of course changes the values of the decimal numbers specifying the function), a variety of circuit arrangements may generally be obtained. Due to the variable criteria for circuit economy, some circuit arrangements may be more economical than others. All circuits will however be good ones; in many cases considerably better than circuits obtained by other methods. The reason for this is that if any symmetry exists, then economical nonseries-parallel circuits automatically evolve.

This paper has concerned itself primarily with the exposition and illustration of this method. Consequently, the examples chosen are of the simplest types, and the exposition has been restricted to the case of the 2-terminal switching circuit.

References

1. Boolean Matrices in the Design of Combinational Switching Circuits, F. E. Hohn, L. R. Schissler. *Bell System Technical Journal*, New York, N. Y., vol. 34, 1955, pp. 177–202.

2. Design of Switching Circuits (book), W. Keister, A. E. Ritchie, S. H. Washburn. D. Van Nostrand Company, Inc., Princeton, N. J., 1951.

3. A Symbolic Method for Synthesis of 2-Terminal Switching Circuits, D. Zeheb, W. P. Caywood. *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 690–93.

4. The Map Method for Synthesis of Combinational Logic Circuits, M. Karnaugh. *Ibid.*, vol. 72, pt. 1, Nov. 1953, pp. 593–99.

Deficiency of Standards for High-Voltage High-Power Rectifier Transformers

JAMES J. HALLORAN
MEMBER AIEE

Synopsis: This paper may be considered as a status report in the development of standards for electronics transformers used with high-vacuum rectifier tubes. The need for adequate standards for this type of transformer has become more pressing as vacuum rectifier tubes of high voltage and high power have been developed to meet the requirements of research and manufacturing operations. Military and commercial applications that are now being developed add to the need for adequate standards.

AN EXAMPLE of the research application is the proton linear accelerator at the University of Minnesota. The resnatron tubes providing radio-frequency power for the accelerator require a 110-kv 4-ampere d-c supply described in a 1953 paper.[1] The plate transformers for this d-c supply consist of three 167-kva single-phase 60-cycle oil-insulated transformers 960 volts primary to 90,000 volts secondary for delta-delta 3-phase full-wave rectifier service. No adequate standards for dielectric testing of these transformers existed at the time of their manufacture, nor do they to this time, and this was noted in one of the printed discussions of the afore-mentioned paper. However, the development of a standard is not accomplished by a single discussion. Many steps were taken in the years that followed to bring together engineers sufficiently interested in this subject to take some action. A brief résumé of these steps, omitting personalities, may be of interest to other engineers seeking a pattern of action in the development of standards for their particular field.

Correspondence with the American Standards Association in 1953 indicated a lack of interest in the electronics transformer field by committees of that organization. Turning to AIEE members for assistance in 1955, the matter was discussed with one of the AIEE Directors who had broad interests in both the communication and power fields. This led to presentation of the problem to the chairman of the Standards Committee of AIEE, who in turn referred the matter to the chairmen of the Transformers Committee, the Science and Electronics Division Committee, and the then Committee on Technical Operations. All of these chairmen expressed their willingness to consider the subject. In June 1956, the matter was presented to the Science and Electronics Division Committee assembled in San Francisco, Calif., during AIEE's Summer General Meeting. This committee took up the problem, and over a period of months reached the conclusion that electronics transformers have been a neglected subject in AIEE and moved to rectify this condition by the formation of an Electronics Transformers Subcommittee within the Electronics Committee of the Science and Electronics Division of AIEE.

ELECTRONICS TRANSFORMERS SUBCOMMITTEE

The subcommittee held its first meeting in New York City on March 20, 1957. The scope of this new subcommittee is as follows:

"Treatment of design, construction, and operation of transformers, reactors, and similar equipment as applied to circuits and systems which include electronic devices as circuit elements or as loads. It is the intent here to treat such devices as are not already covered by other committees."

The first group of papers sponsored by this subcommittee was presented at the AIEE Fall General Meeting in Chicago, Ill., in October 1957.

Present Standards

Standards most often applied to electronics transformers by manufacturers are ASA Standard *C 57* for Transformers, Regulators and Reactors and RETMA Standard *TR-102-B* for Power Transformers for Radio Transmitters.[2,3] One of the most interesting comparisons to be made between these standards is in their respective low-frequency dielectric tests.

ASA STANDARDS

ASA C57.12 (Distribution, Power, and Regulating Transformers...) Table 12.060. Typical tests given in this table are 50-kv rms for an oil-filled transformer of the 25-kv insulation class, 140-kv test for the 69-kv class, and 325-kv test for the 161-kv class. The insulation class of 3-phase transformers is equal to the line-to-line voltage. Thus the test voltage is approximately two times the line voltage. This is considered adequate to test the unit for its ability to withstand stresses occurring when one line is grounded. The adequacy of these tests has been proved by the experience of the public utilities over many years.

ASA C57.18 (Rectifier Transformer Equipment used with Pool Cathode Mercury-Arc Power Converters). Paragraph 18.040 (b) reads: "The d-c windings of rectifier transformer equipment shall receive low frequency tests in accordance with...Table 12.060. The insulation class shall be determined by the square root of three divided by the square-root of two times the highest operating crest voltage between terminals and case...the highest crest voltages to ground are indicated in Table 18.040." For a 3-phase full-wave rectifier, Table 18.040 gives a maximum crest voltage of the square root of six times the voltage to neutral. Thus for a line-to-line voltage of 25 kv, the highest operating crest voltage is 25 kv divided by the square-root of three times the square root of six or the familiar square root of two times 25-kv or 35-kv peak. From this, the insulation class is calculated to be 42.5 kv. Since intermediate classes take the test of the next higher class, the required test becomes 95-kv rms in accordance with the 46-kv class, or a test value almost twice as high as that required for 25-kv power transformers. It may be argued that *C57.18* is not intended to apply to high-voltage rectifier transformers but customers specifications often follow these standards in lieu of a better standard.

RETMA STANDARDS

RETMA *TR-102-B* lists dielectric strength tests as follows:

"1. A test potential of twice the rms working voltage plus 1,000 volts shall be applied between each winding and each other winding and between each winding and the core or case. All windings not under test shall be grounded to the core or case.

NOTE: Since the application of test

Paper 57-1014, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7-11, 1957. Manuscript submitted July 8, 1957; made available for printing August 19, 1957.

JAMES J. HALLORAN is with Electro Engineering Works, San Leandro, Calif.

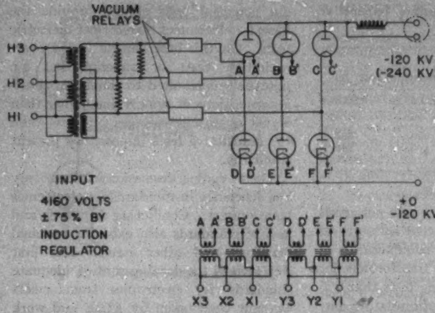


Fig. 1. Schematic diagram of 120-kv 2-ampere d-c supply

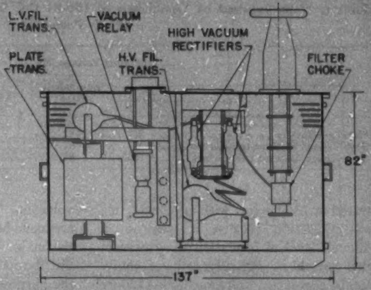


Fig. 2. Cross-sectional drawing of 120-kv d-c supply showing location of components. Low-voltage components tested at 185-kv rms 60 cycles, one minute. High-voltage components tested at 300-kv rms 60 cycles, one minute, with low-voltage components at 150-kv rms simultaneously

potentials may impair the insulation, any dielectric strength test shall, if repeated, be made at not more than 90% of the specified test potential.

2. The test potentials shall be increased gradually, i.e., at the rate of approximately 2 kv per second, from zero to the specified values, maintained at the specified values for a period of one minute, and decreased gradually to zero."

Thus a 25-kv rms transformer would be tested at 51-kv rms which is in close agreement with the ASA power transformer standards.

Applying Present Standards

It will be of interest to compare these dielectric test standards in a typical assembly of plate and filament transformers and high-vacuum rectifier tubes. Fig. 1 is a schematic of a d-c power supply used in the manufacture of klystron tubes. This supply may be operated with the positive terminal at ground potential or in cascade with another supply at 120 kv below ground potential. In the first case the plate transformer has a maximum line-to-line voltage of 97-kv rms and a peak voltage of 1.414 times 97-kv or −137-kv peak. In the second case, the peak voltage will be −137–, −120–, or −257–kv peak. Tests required by the three foregoing standards are indicated in Table I.

The high-voltage filament transformers are connected to the plate transformer terminals and are subjected to the same voltages as above. Their winding voltages (20 volts) contribute no significant value to the operating voltage. The low-voltage filament transformers operate at either ground potential or −120-kv peak.

Curves showing creep distances over phenolic resin impregnated materials in air and in oil indicate the compounding of problems above 400-kv peak. By the time 500-kv peak is reached, oil clearances have no advantage over air clearances. Transformer design circumvents this as far as possible by using solid insulation in place of creepage distance. However, connections to tubes, relays, and resistors present problems which can be eased by a more realistic appraisal of test voltages. Furthermore, it is recognized that alternating potentials have a deteriorating effect on insulation due to heat developing in the insulation as it is charged and discharged. RETMA standards recognize the destructive nature of alternating potential tests by reducing test voltages subsequent to the first test. Thus, test values of alternating potentials should be held to a minimum consistent with proving a good factor of safety in the product.

On the plate and high-voltage filament transformers the 182-kv rms "equivalent" to the −257-kv peak voltage was considered to be in the 161-kv class instead of in the next class above it (196-kv class, 395-kv test) and a 325-kv test was considered. This test was further reduced because of the fact that −120 kv o the peak voltage rating is actually d-c potential which does not deteriorate the insulation to the same extent as an alternating potential. Final test voltages were established at 300-kv rms. The low-voltage filament transformers were

Fig. 3. Installed 120-kv 2-ampere d-c supply. Unit had operated 700 hours without a rectifier tube replacement or other component service at the time of this photograph

Table I. Tests Required on Transformers of 120-Kv D-C Supply with Positive Terminal at −120-Kv Potential

	Operating Voltage, Kv		Low-Frequency Dielectric Test, Kv		
	Peak	"Equivalent" Rms	ASA C57.12	ASA C57.18	RETMA
Plate transformer	−257	182	395	690	365
High-voltage filament transformer	−257	182	395	690	365
Low-voltage filament transformer	−120	85	185	325	171

tested at 185-kv rms without considering any reduction because it was not a difficult level of insulation to achieve in this assembly. Actually, units of equivalent operating voltage level (120-kv direct current) have been in operation satisfactorily for years after having been tested at one and one half times "equivalent" rms voltage instead of double "equivalent" rms voltage; see Figs. 2 and 3.

The term "equivalent" rms voltage has been used parenthetically in the foregoing discussion because no rms test is actually equivalent to a d-c or unidirectional potential test. It is very easy to over-engineer a unit simply to meet rms test voltages that have only a limited relation to the actual operating voltages. The use of d-c test potentials are now being investigated by some of the major transformer manufacturers of distribution and power transformers for the purpose of developing tests that will avoid the deterioration of insulation during testing. This new philosophy is even more pertinent to electronic transformers where units are operating at direct or combined direct and alternating voltages.

Conclusions

In the development of standards for electronics transformers, it is the author's opinion that a clean break should be made away from a-c potential tests and that extensive research into the use of d-c potential tests should provide the basis for electronics transformer dielectric tests. On this base can be added overexcitation tests or superimposed a-c potentials as required to simulate operating conditions at some value higher than actual operating voltages but less than the destructive tests imposed by present standards.

The foregoing discussion explores only one deficiency in standards for electronics transformers. Conflicting standards and lack of standards also exist for terminal markings and other items. The first step toward the development of adequate standards for electronics transformers has now been taken by AIEE and work must continue rapidly to keep up with the pace of military and commercial demands.

References

1. Oil-Immersed 110-Kv 440-Kw Rectifier Unit for Resnatron Plate Supply, R. P. Featherstone, C. Ling, P. A. Cartwright. *AIEE Transactions*, vol. 72, pt. I, May 1953, pp. 145–52.

2. Transformers, Regulators and Reactors. *ASA Standard C 57*, American Standards Association, New York, N. Y., 1948.

3. Power Transformers for Radio Transmitters. *RETMA Standard TR-102-B*, Radio, Electronics and Television Manufacturers Association, Washington, D. C., 1950.

◆

Discussion

S. T. Maunder (General Electric Company, Holyoke, Mass.): Mr. Halloran has very properly brought to the attention of the Institute the lack of standards for high-voltage high-power rectifier transformers for electronic service. His paper calls for the proper development of a standard that should be a clean break away from a-c potential tests to d-c potential tests. Most of the transformer art and practice has been developed based on power utility usage. The specifications that Mr. Halloran cites, reference 2 of the paper, were written for this type of equipment. The surge voltages on the utilities' lines are high; even so the utilities have found it economical to obtain transformers designed and tested to stand twice normal voltage, and in addition to protect the units with arresters.

The higher voltage electronic circuit art and practice is relatively new. Insufficient work has been done to determine the surge voltages in the circuits. This area offers a fruitful field of study. Until more is known therefore, the application of the correct insulation level of a transformer to a circuit must be the responsibility of the user for only he can determine or limit the surge voltages in his circuits.

The transformer insulation can be tested in three well-known ways:

1. Applied voltage test:

This is the well-known a-c high-potential test. It subjects the major insulation structure, the leads, and the bushings to overpotential to assure proper margin over and above normal voltage. In general terms this is usually twice normal voltage, a-c. Mr. Halloran suggests the use of a d-c test for rectifier transformers. An examination of the voltage on the various areas of the transformer insulation will show that the voltage stresses are a-c. Further, the voltage stress is distributed slightly differently with alternating current (by capacitance) than by direct current (by insulation resistance). For insulation structures like transformers, d-c test voltage searches out d-c insulation weaknesses while a-c test voltage searches out a-c insulation weaknesses. Neither searches out the significant weaknesses of the other voltage. Thus it is more proper to test with alternating current to simulate the voltage distribution more exactly in actual use.

2. Induced voltage test:

While the applied voltage test stresses the major insulation, yet the induced voltage test is used to stress turn-to-turn, or coil-to-coil insulation structures. Further, it is the accepted test for units which have one end grounded; thus effectively testing the major line insulation as well as the turn to turn or coil to coil. Usually this is at twice normal voltage and at a higher frequency to avoid overexcitation of the core.

3. Impulse voltage test:

Introduced in the last generation, this test supplements the others mentioned and permits a test of the insulation structure and turns by an impulse wave. Co-ordination of the known protective equipment on the transformer with the impulse-tested insulation structure can assure a user of the ability to withstand estimated lightning surges on a transmission line. Presently there is little need of this type of test for electronic transformers until more is known of the service conditions to which the transformer is subjected.

G. Leslie Hill (Hill Research Company, Oakland, Calif.): This paper is of particular interest as it points out the need for adequate test standards for electronic transformers operating on combined direct and alternating voltages.

The author points out that "it is recognized that alternating potentials have a deteriorating effect on insulation due to heat developing in the insulation as it is charged and discharged." The deteriorating effect of high a-c test potentials has long been recognized in the power industry.

Experience, engineering test data, and research[1-3] have shown that both a-c acceptance tests and maintenance overvoltage tests can damage the insulation without the knowledge of the test engineer. Puncture of the insulation may occur a few seconds prior to and/or at the termination of the one-minute test period. Puncture can also occur after the one-minute test period and after the operator has started to reduce the alternating voltage by use of the regulator. It is generally accepted that the a-c high-potential test is a "go" or "no-go" test, and no method of predicting the dielec-

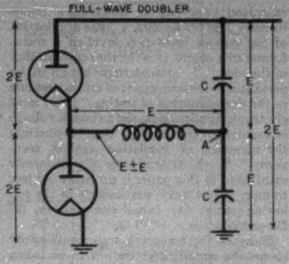


Fig. 4. Full-wave voltage doubler

tric strength of the insulation without puncturing it has heretofore been found.

The d-c potential tests on electronic transformers would be of especial value where the transformers are to supply the plate and filament potentials in high-voltage rectifiers used for supplying d-c potentials in the measurement of the current and voltage values used in the diagnosis of electrical insulation.

It is obvious that a very high resistance between the primary and secondary is necessary to avoid flow of direct current back to the grounded leg of the primary circuit and in turn back through the microammeter in the ground return of the high-voltage direct current. A reading of a few microamperes will cause an appreciable error when testing insulation having a high insulation resistance.

It also must be recognized that any a-c overpotential test will not give any information of the d-c resistance of the insulation except to prove it has passed the a-c overpotential tests.

The author is to be commended for his recommendation of the use of high-voltage d-c potential tests on electronic transformers. It will enable the design engineer to diagnose the insulation he is using.

References

1. Alternating and Direct Voltage Endurance Studies on Mica Insulation for Electric Machinery, Graham Lee Moses. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 763–69.

2. Testing Electrical Insulation of Rotating Machinery with High-Voltage Direct Current, G. Leslie Hill. *Ibid.*, vol. 72, pt. III, Apr. 1953, pp. 159–74.

3. Diagnoses of A-C Generator Insulation Condition by Nondestructive Tests, A. W. W. Cameron. *Ibid.*, vol. 71, pt. III, Jan. 1952, pp. 263–74.

Hugh C. Ross (Jennings Radio Manufacturing Corporation, San Jose, Calif.): In view of the tremendous increase in the number of and the power of the higher power transmitters and other types of so-called electronic types of loads, this paper and the action it suggests taking is very timely.

In comparison to many other types of transformers, the general type of transformer mentioned in the paper is probably subject to more severe duty, particularly in the following areas.

1. Initial energization: In this use, energization and de-energization may occur very frequently; therefore, insulation may be stressed more than usual by the forces occurring from the repeated high inrush or charging currents each time an unmodified primary contactor is closed. Also, these inrush currents, possibly multiplied by any bounce existing in the contactor, may create serious transient oscillations which can create excessive transient overvoltages, particularly in the end turns and possibly in the entire secondary windings.

2. De-energization:
(*a*) Under no load.
(*b*) Under normal load.
(*c*) Under overload or fault.

Frequent de-energization without proper switching methods may create serious transient overvoltages on either primary or secondary windings. Frequent overload or fault conditions may again add excessive or repeated mechanical stress on insulation. Yet another possible insulation problem is created by the newly developed high-speed vacuum circuit breakers which can interrupt in one-half cycle or less, depending on contact parting speed, shunt capacitance and current. It has been found that if flashovers in high-power tubes can be interrupted in less than the time of one cycle, the tube is much less likely to be damaged by a power arc and may be able to sustain a great many of these short-time flashovers. Therefore, many power supply users are requesting the use of this high-speed interruption even though the sudden interruption of faults, especially during resulting asymmetric current conditions, may require more severe transient voltage duty or suppression during the resultant rapid field collapse.

It would appear that many of the transformer failures may then occur because of these transient overvoltages rather than continuous 60-cycle voltage. This would indicate that some type of impulse test voltage, if properly applied and interpreted, would possibly be much more revealing for this type of transformer than the present and probably damaging 60-cycle overvoltage tests now being used. It would be interesting to know if this type of test is being considered as part of the over-all testing program. Also, can there be a less severe test used where transient suppression, such as nonlinear resistors, arresters, or small shunt capacitors, and proper end-turn protection is incorporated?

From experience in the use of d-c high-potential in conjunction with a-c high-potential tests on transformer insulation, it is surprising that the use of the less destructive and more indicative direct current is not more widely demanded. Extremely favorable results have been obtained with d-c leakage checks where a-c overvoltages either were nonindicative or damaging, particularly in periodic maintenance checks or during manufacture. Is there any definite plan to apply this test in the testing program?

It is interesting to note that this paper mentions the first complete high-power supply with built-in 1/2-cycle vacuum circuit breakers. The unit including the vacuum switches is completely immersed in oil and withstood over 2,000 energizing and de-energizing operations in the first 700 hours of operation. Most of the de-energizing operations were 1/2-cycle fault interruptions by the 100-kv rms 3-phase instantaneous vacuum circuit breaker in the high-voltage secondary, although some were by the moderate-overload inverse-time-delay vacuum circuit breaker in the 4,160-volt primary side.

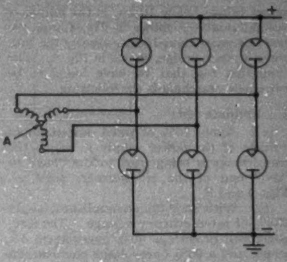


Fig. 5. Full-wave 3-phase bridge

Is it possible that these new standards may indicate or solve some of the many problems that arise in the type of power supply using these transformers? It has long been felt that the combined experience and knowledge of the user, *and* the transformer manufacturer, *and* the switchgear manufacturer, should be used together in each of these applications. It has been found all too often in the past that when the determination of insulation or testing specifications has been left entirely up to the user, that the result was either unnecessary safety factors and resultant excessive cost, size, or weight, or insufficient safety factors and resultant early failure and user dissatisfaction. It would appear that it would be necessary, if not mandatory, that all major participants in the completed unit contribute in the determination of testing or safety specifications to ensure a reliable and reasonable unit.

Victor Wouk (βeta Electric Corporation, New York, N. Y.): Mr. Halloran has called vividly to the attention of our profession, facts concerning the unrealistic nature of low-frequency tests' standards in regard to dielectric strength, when applied to high-voltage high-power rectifier transformers. I would like to extend the comments to low-power high-voltage transformers, i.e., high-voltage transformers for rectifiers in electronic applications.

The βeta Electric Corporation has been specializing in the design and construction of high-voltage d-c power supplies since 1946, from voltages of 10,000 volts to over 500,000 volts, at currents ranging from 1 milliampere to 10 amperes. The "plate transformers," or rectifier transformers, have been built to pass specifications or meet standards considerably less stringent than ASA *C57.12*, etc. We have built and installed more than 4,000 power supplies, with a reliability record that is completely satisfactory to the industry. The designs have been adopted by most manufacturers of transformers for high-voltage d-c power supplies in the electronic industries.

The circuit employed by most designers of high-voltage d-c power supplies, in the power ranges below 10 kw, is the full-wave voltage doubler,[1] shown in Fig. 4, and, for power greater than 10 kw, usually the full-wave 3-phase bridge, shown in Fig. 5. The insulation test that we have found to be extremely searching is the following.

1. 1-minute test:

(*a*). Apply 70% of rated output direct voltage of the power supply in which the transformer is being used to the "low" or "d-c" end of the transformer, point *A*, Figs. 4 and 5.

(*b*). With step (*a*) accomplished, apply 135% of rated primary voltage. The foregoing will produce a 40% overvoltage d-c stress, and a 90% overvoltage combination a-c–d-c, i.e., peak stress. Acceptance is two or fewer flashovers, and no insulation puncture. More than two flashovers, or insulation puncture, will cause rejection.

2. 12-hour test:

(*a*). Apply 60% of rated output direct voltage of the power supply in which the transformer is being used to the "low" or "d-c" end of the transformer, point *A*, Figs. 4 and 5.

(*b*). With step (*a*) accomplished, apply 120% of rated primary voltage.

The foregoing will produce a 20% overvoltage d-c stress, and a 44% overvoltage combination a-c–d-c, i.e., peak stress. Acceptance is two or fewer flashovers, and no insulation puncture. More than two flashovers, or insulation puncture, will cause rejection.

We have found, with these tests, less than 1/2% field failures of plate transformers. We believe that the reasons for the adequacy of these tests, in comparison with the much more severe existing standards that would cause electronic applications' high-voltage d-c power supplies to cost substantially more than they do now, without an equivalent increase in reliability, are mainly the following.

1. Lack of secondary transients: In most electronic applications, the primary voltage is usually applied slowly, from zero, or a fraction of full-line voltage. The full-line voltage is not applied suddenly, as the use of a COVAT (Continuously Variable Auto-Transformer, i.e., variac, powerstat, etc.) is almost universal, thus eliminating the necessity of insulating against transients that can be as great as twice rated secondary voltage.

2. Less stringent duty cycle: In most electronic applications, the power supply is rarely used at full rated output voltage for long periods of time, so that the duty cycle is low. Furthermore, the power supply is usually employed at output voltages well below its maximum capabilities, in tube testing, developmental and experimental projects, etc. Therefore, the test procedures outlined above may be as stringent, relative to the usual actual operating voltages, as is required by existing standards.

3. Greater d-c strength: It is well known that solid dielectric materials can have long-time dielectric strengths many times greater for d-c applied voltages, than are experienced with alternating voltages.[2] Therefore, considerably lower margins of safety are required for insulating materials withstanding d-c stresses than the same materials under a-c stresses.

βeta has designed and built filament transformers that withstand 300-kv d-c successfully in 24 hour year-round duty, which would fail rapidly if tested at 80 kv a-c.

We at βeta Electric hope that a more formal program will be undertaken by the AIEE members interested in the subject of electronic transformers, to adopt more realistic test standards, as recognized by Mr. Halloran. Our experience can be used as a basis for evaluation of recommended new standards.

REFERENCES

1. THE FULL-WAVE VOLTAGE DOUBLING RECTIFIER CIRCUIT, D. L. Waldelich. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 29, 1941, p. 554.

2. CIRCUIT COMPONENTS FOR HIGH VOLTAGE DC POWER SUPPLIES, V. Wouk. *WESCON Convention Record*, session 10, pt. 6, 1957.

J. J. Halloran: The discussions present several points of view relative to high-potential testing of transformers used in electronic circuits, and the divergence of these views emphasizes further the need for a forum of discussion which has only recently become available by formation of the Electronics Transformers Subcommittee. The quoting of a few phrases from these discussions will emphasize the differences of opinion and the wide areas which must be examined to set up an Electronics Transformers Test Standard: ". . the application of the correct insulation level of a transformer to a circuit must be the responsibility of the user." ". . mandatory that all major participants in the completed unit contribute in the determination of testing." "In most electronic applications, the primary voltage is usually applied slowly, thus eliminating the necessity of insulating against transients." "..the general type of transformer mentioned in this paper is probably subject to more severe duty, particularly in the following areas: (1) Initial energization, (2) De-energization. . ."

Some of the foregoing statements apply to specific circuits and are not to be taken in a generalized sense. However, in order to write a standard of general use, it will be necessary to establish insulation tests which apply to the transformer, irrespective of the connected circuit, and then recognize variances for specific applications.

Impulse voltage tests are mentioned by Mr. Maunder and Mr. Ross, and ASA standards *C57.12* list impulse voltage values which most transformers will withstand. However, these are destructive tests and shorten the life of insulation even when they do not destroy it. It would be desirable to replace these impulse tests by milder pulse voltages combined with improved corona-measuring techniques.

The combined a-c–d-c tests used by Mr. Wouk are excellent tests and, if combined with measurements of direct current flowing through the insulation as described in references listed by Mr. Hill, will probably tell not only if the insulation is good, but also how good. However, having once established the combined a-c–d-c characteristics of insulation in a transformer, d-c tests only may be used more economically and without damage to the insulation. This line of reasoning is justified on the same basis as initial temperature rise tests are replaced by simpler excitation and impedance tests once the design is proved. As a better understanding of a-c–d-c characteristics of insulations is obtained, it is not improbable that basic reference data on insulations combined with quality control of their use in manufacture will permit evaluation of insulation dielectric strength in the product by d-c tests alone. This is the end result toward which the paper points.

A Generalized Theory of Transistor Bias Circuits

HERBERT HELLERMAN
ASSOCIATE MEMBER AIEE

ONE of the important problems met in the design of transistor circuits is how to arrange the bias network to obtain the proper operating point and how to maintain this condition within specified limits, subject to environmental changes such as temperature, or to a range of parameters due to production spread in device manufacture. Several bias circuits have been proposed and are being used which give good performance. The question naturally arises as to what essential features do these circuits have in common? It is the purpose of this paper to introduce a very general method

---

Paper 57-1023, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted December 6, 1956; made available for printing August 14, 1957.

HERBERT HELLERMAN is with the University of Delaware, Newark, Del., formerly of Syracuse University, Syracuse, N. Y.

The author wishes to thank Professor F. Reza for helpful discussions on this subject which led to the study reported here. Portions of this work were performed as part of a contract between Electronic Associates, Inc., and Syracuse University Research Institute.

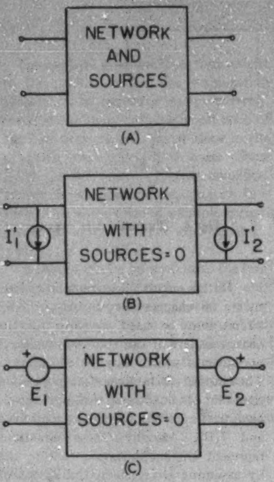


Fig. 1. Equivalent circuits for networks containing sources

of analyzing bias circuits and begins by first showing that almost all 3-terminal bias circuits can be reduced to the same simple standard form. The analysis for this standard-form circuit will then be presented. Finally, an example drawn from current practice will be given showing the unifying concepts which result from the general theory.

The Generalized Thevenin-Norton Theorems

The analytical problem of determining the bias supplied to a 3-terminal amplifying device usually reduces to solving for the currents and voltages at the device when it is connected to a network consisting of some arrangement of d-c sources and resistors. From a network point of view, the bias circuit external to the device is a 3-terminal linear network containing sources. The first step is then to find a general equivalent circuit of such a network. It is readily proved (see appendix) that the circuits of Figs. 1(B) and 1(C) are equivalent to that of Fig. 1(A) with respect to the network terminals. Fig. 1(B) gives the equivalence in terms of a passive network and two current generators. This may be considered a generalization of the familiar Norton theorem of two terminal networks. The passive network consists of the original network with all sources zero. The generators are the short-circuit currents at the respective terminals. Fig. 1(C) expresses the equivalence in terms of a passive network and two series-voltage generators which may be determined by open-circuit measurements at the terminals. These theorems are also useful in connection with the noise theory of 4-terminal networks.[1] Although the emphasis here is on transistor bias circuits the principles are easily applied to vacuum tube or any other device circuit.

Bias Circuit Analysis

The application of the theorem to the bias circuit problem is shown in Fig. 2. The equivalent circuit of Fig. 2(B) uses a *T* representation for the passive network (although, of course, this is but one of several possible generalized configurations). The elements of the *T* are expressed in terms of the *Z* parameters of the bias circuit. It is evident that this structure represents the most general case since any circuit may be reduced to this form. For this reason, it will be called the standard form circuit.

Writing the loop equations and Kirchhoff's current law

$$v_{BE}-i_E r_{12}+i_B(r_{11}-r_{12})-E_1=0$$

$$v_{CE}-i_E r_{12}+i_C(r_{22}-r_{12})-E_2=0 \qquad (1)$$

$$i_E+i_B+i_C=0$$

the third equation may be used to eliminate i_E from the other two, giving

$$v_{BE}+i_B r_{11}+i_C r_{12}-E_1=0 \qquad (2A)$$

$$v_{CE}+i_B r_{12}+i_C r_{22}-E_2=0 \qquad (2B)$$

There are four unknowns (v_{BE}, v_{CE}, i_B, i_C) indicating that two more relations are needed among the variables.

Graphical Solutions

The most exact solution is obtained by giving the two needed relations in the form of two sets of static curves. One of these can be the familiar collector characteristics which in its most useful form plots i_C versus v_{CE} with i_B as the parameter. The other set shows v_{BE} versus i_B with i_C as the parameter. An orderly method for obtaining the operating point begins by plotting equation 2(B) on the collector characteristics with i_B assumed fixed say at I_{B1}. This is evidently a straight line which will intersect the I_{B1} curve of the collector characteristic at a point v_{CE1}, i_{C1}. Another line corresponding to equation 2(B) with $i_B=I_{B2}$ will, in a similar manner, give an intersection with i_{B2} of the collector characteristics at coordinates v_{CE2}, i_{C2}, i_{B2}, etc. A curve connecting these points can be drawn which then satisfies equation 2(B) and the transistor collector characteristics simultaneously. The operating point certainly lies somewhere on this curve. The next step is to perform the same construction as outlined above using equation 2(A) and the input characteristics of the transistor. By direct analogy with the above discussion, the result is a curve in the v_{BE}, i_B, i_C plane. Now points read off this curve may be plotted as a new curve on the collector characteristics. The intersection of this curve with the one already there locates the operating point.

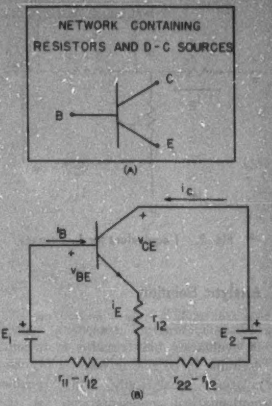


Fig. 2. General transistor bias circuits

Although the foregoing process can give a good solution to the problem, it requires two sets of curves, only one of which is usually supplied by the device manufacturers. Furthermore, the method is somewhat cumbersome and it is difficult to derive general results from it. Considerable simplification results if the assumption $v_{BE}=0$ is made. This is normally a good approximation and it is a standard one often used in bias calculations.[2] Equations 2 then become

$$i_C=-v_{CE}\frac{r_{11}}{\Delta}+\left(E_2\frac{r_{11}}{\Delta}-E_1\frac{r_{12}}{\Delta}\right) \qquad (3)$$

where

$$\Delta=r_{11}r_{22}-r_{12}^2$$

Equation 3 is a load-line type of relation on the collector characteristics. The operating point lies on this line and may be found by its intersection with the curve resulting from the plot of equation 2(A) with $v_{BE}=0$. This method has the advantage of requiring but a single set of curves but is approximate.

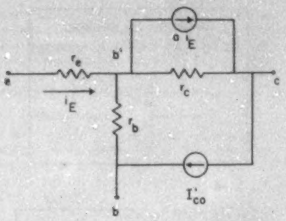


Fig. 3. T equivalent for I_{co} changes

Analytic Solution

In order to arrive at a general method of evaluating bias circuits, it is most desirable to be able to perform bias calculations analytically. This requires mathematical representations for the static curves which can then be used with equations 2. In most cases this can be done to a usable approximation by idealizing the collector characteristics to straight, equally spaced lines, at least beyond the region of very low collector voltages. Then

$$i_C = \alpha_{fe} i_B + I_{CeO} \tag{4}$$

where

α_{fe} = signal forward short-circuit current amplification in the common emitter configuration
I_{CeO} = collector current which flows when the base is open and collector is normally biased

For completeness, the relations between α_{fe}, I_{CBO}, and the familiar transistor parameters α (or α_{fb}) and I_{CO} (or I_{CBO}) are

$$\alpha_{fe} = \frac{\alpha}{1-\alpha} \tag{5}$$

$$I_{CBO} = I_{CO}\alpha_{fe} \tag{6}$$

I_{CO} and hence I_{CBO} can be considered to be made up of two parts.[3] One of these I'_{CO}, is essentially independent of voltage and is due entirely to diffusion effects in the transistor. The other part is due to ohmic leakage and is proportional to collector voltage. Thus

$$I_{CO} = I_{CO}' + \frac{v_{ce}h_{oe}}{\alpha_{fe}} \tag{7A}$$

$$I_{CeO} = I_{CeO}' + v_{CE}h_{Oe} \tag{7B}$$

This division of I_{CO} or I_{CeO} is of importance because the parts independent of voltage are very temperature sensitive, as will be discussed later.

Using equations 2, 4, 6, and 7(B), the solutions for collector current and voltage in any bias circuit are, with a little algebra, found to be

$$i_C = \frac{\alpha_{fe}I_{CO}'r_{11} + E_1(\alpha_{fe} + r_{12}h_{oe}) + E_2r_{11}h_{oe}}{r_{11} + \alpha_{fe}r_{12} + h_{oe}\Delta} \tag{8}$$

$$v_{CE} = -\frac{\alpha_{fe}I_{CO}'\Delta - E_1(\alpha_{fe}r_{22} + r_{12}) + E_2(r_{11} + \alpha_{fe}r_{12})}{r_{11} + \alpha_{fe}r_{12} + h_{oe}\Delta} \tag{9}$$

Since I_{CO}' cannot be measured directly, while I_{CO} can be determined by a simple measurement, it is of interest to note that equations 8 and 9 may be written with the primes removed if h_{oe} is set equal to zero, giving

$$i_C = \frac{\alpha_{fe}I_{CO}r_{11} + E_1\alpha_{fe}}{r_{11} + \alpha_{fe}r_{12}} \tag{10}$$

$$v_{CE} = E_2 - \frac{\alpha_{fe}I_{CO}\Delta + E_1(\alpha_{fe}r_{22} + r_{12})}{r_{11} + \alpha_{fe}r_{12}} \tag{11}$$

Bias Stabilization

The design of transistor bias circuits is complicated by the fact that the currents which flow in the transistor are temperature sensitive. A common requirement is that the circuit be so designed that the transistor remains biased to provide useful amplification over a specified range of temperatures. In most cases the amplifier is only required to amplify signals without reproducing their d-c levels. In such cases the degenerative effects of the circuits used for bias stabilization may be eliminated as far as the signal is concerned by appropriate by-passing. The process of bias stabilization is fundamentally one of the application of degenerative d-c feedback as will be demonstrated using the theory developed earlier.

One of the principal sources of temperature variation of bias point is the change of I_{CO}' with temperature. The variation follows an exponential law in all semiconductor junction devices. This may be described for germanium units by saying that I_{CO}' doubles for every 10 C change in temperature. The value of I_{CO}' at room temperature in typical low-power germanium transistors is between 1 and 15 microamperes. Although the same type law is followed in silicon devices, I_{CO}' is much smaller at any given temperature than in germanium units.

If only variations in I_{CO}' are of interest as in the bias stabilization problem then one can handle changes in the currents of the circuit as signal currents. The problem may then be solved using the methods of signal analysis, e.g., the transistor equivalent circuit. The source of drift which is $\Delta I_{OC}'$ then appears as a generator as shown in Fig. 3. Actually little error is introduced if the base end of this generator is considered to be at b'. In this circuit all currents are increments or changes due to $\Delta I_{CO}'$. This circuit may be used to solve for the change in any of the transistor voltages or currents which may be of interest. The solution can be somewhat more exact than that obtained using the approximate analytical approach based on equations 4 and 7(B). Actually, these equations represent the simplification that results by assuming $r_e = r_b = 0$ in the circuit of Fig. 3. However, although more general, the use of the incremental equivalent circuit is complicated by the limitations of using such a representation. Among these is the fact that the resistances in the equivalent circuit are all temperature sensitive as is the parameter a. Furthermore, the parameters are dependent upon the bias conditions especially the collector current. Fortunately, all these are usually second order effects in the ordinary temperature ranges of interest. In many cases the highly simplified representation of equation 4 suffices to obtain the results desired. The one case where this is likely to lead to appreciable error is when the d-c load impedance on the transistor is high. Even here the most general circuit of Fig. 3 is not essential although r_c or equation 7(B) should be included in the calculations.

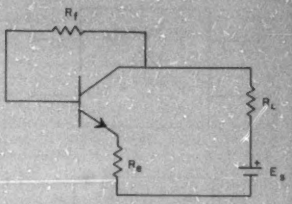


Fig. 4. Illustrative bias circuit

Two figures of merit which are often used to evaluate bias circuits with regard so temperature changes are the current stability factor[4]

$$S_I = \frac{\partial i_c}{\partial I_{co}} \tag{12}$$

and the voltage stability factor

$$S_V = \frac{\partial v_{CE}}{\partial I_{CO}} \tag{13}$$

These quantities are easily found for the general standard form circuit from equations 10 and 11.

$$S_I = \frac{\alpha_{fe}}{1 + \alpha_{fe}\dfrac{r_{12}}{r_{11}}} \tag{14}$$

$$S_V = \frac{\Delta}{r_{11}} S_I \tag{15}$$

Equation 14 gives an important insight to the nature of the bias stabilization process. The quantity r_{12}/r_{11} is recognized as the short-circuit current ratio of the bias network considered alone in the direction from collector to base. Since α_{fe} is the short-circuit current amplification of the transistor alone, equation 15 is seen to be of the form of a simple current feedback equation. The equivalent input drift is then ΔI_{CO}.

It should not be inferred from the foregoing discussion that I_{CO} is the only temperature sensitive quantity. However, it is usually the most important and was used to illustrate the application of the general theory. Since the purpose of this paper is primarily to present a unified method of analysis and evaluation no attempt will be made to give design formulas.

Example

The theory will now be applied to the bias arrangement shown in Fig. 4. The parameters of the bias circuit are easily determined to be

$$r_{11} = R_e + R_f + R_L \qquad \Delta = R_f(R_L + R_e)$$
$$r_{12} = R_L + R_e \qquad E_1 = E_s$$
$$r_{22} = R_L + R_e \qquad E_2 = E_s$$

The collector to emitter bias voltage is then found by substitution into equation 11 and simplification to be

$$v_{CE} = \frac{R_f[E_s - I_{co}\alpha_{fe}(R_L + R_e)]}{R_f + (R_e + R_L)(1 + \alpha_{fe})}$$

The stability factor S_I is found from equation 13

$$S_I = \frac{\alpha_{fe}}{1 + \alpha_{fe}\dfrac{(R_L + R_e)}{R_L + R_e + R_f}}$$

The stability factor S_V is

$$S_V = -\frac{\alpha_{fe}(R_L + R_e)}{1 + (\alpha_{fe} + 1)\dfrac{R_L + R_e}{R_f}}$$

It is instructive to note that as far as bias stabilization is concerned, R_L and R_e are in series so that the sum of these two quantities is the essential property of interest rather than their individual values. This is true because the bias stabilization properties of the circuit are independent of where the signal enters or exits or the positions of the power supplies. Thus, in Fig. 4 the same degree of stabilization can result for $R_L = 0$ and R_e some finite value, as for $R_e = 0$ and R_L nonzero. The two cases are identical as can easily be seen by a simple redrawing of the circuit. In an resistance-capacitance coupled stage, since R_L is present as the usually large collector bias resistor R_e can be made zero. Furthermore, for this case the low a-c load resistance (not R_L) which is the input impedance to the next stage often results in but a small amount of loss of signal gain due to the degenerative action of R_f. In those cases where even this cannot be tolerated a bypass of the signal path through R_f as described in reference 3 can be employed to recover all the available gain. Of course, in those cases where R_L has a high d-c value as may occur in an output stage (nontransformer coupled) the bypass arrangement should be used.

When an output transformer is used, bias stabilization is usually accomplished through R_e, the base being returned to the supply through a divider network. In this case, R_e must be bypassed in order to avoid signal degeneration. The size of the necessary bypass capacitor C_e depends upon the lowest frequency, f_0 the amplifier will be called upon to handle, and the signal source impedance Z_g. A useful expression which C_e should satisfy to prevent signal loss is

$$C_e >> \frac{\alpha_{fe}}{2\pi f_0 |Z_g|}$$

Conclusions

The analysis of bias circuits can be greatly systematized through the use of the multiterminal Thevenin theorem to represent the circuit external to the device being biased. The method is limited to linear bias circuits. It also cannot be used to directly calculate the power lost in the bias network since only terminal conditions, and not necessarily the power conditions within the bias circuit, are preserved in the equivalent circuit.

A complete analysis has been made of the 3-terminal bias network which is directly applicable to vacuum tube and transistor circuits. Attention was centered on the transistor case as a timely example of the unifying concepts which result from the theory. Here it was found that the temperature stabilizing properties of the bias circuit can be expressed in the form of a simple feedback equation.

Appendix

The generalized Norton Theorem may be derived by considering a general network containing generators and having a total of M loops. N terminal pairs are brought out of this network. The loop equations of such a system would then be

$$\sum_{n=1}^{M} I_n z_{1n} = E_1 + E_1'$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot$$
$$\sum_{n=1}^{M} I_n z_{Nn} = E_N + E_N' \tag{16}$$
$$\cdot \quad \cdot \quad \cdot$$
$$\cdot \quad \cdot \quad \cdot$$
$$\cdot \quad \cdot \quad \cdot$$
$$\sum_{n=1}^{M} I_n z_{Mn} = E_M'$$

For convenience the first N loops have been taken to be the N external ones. The primed quantities denote the sum of the generator voltages around the respective loops while the unprimed voltages indicate the terminal potentials. The unprimed voltages of course only appear in the first N equations.

The solution of equations 16 by the standard method of determinants yields for the kth current

$$I_k = \sum_{l=1}^{N} E_l \frac{\Delta_{lk}}{\Delta} + \sum_{l=2}^{M} E' \frac{\Delta_{lk}}{\Delta} \tag{17}$$

where Δ is the network determinant on a loop basis and the Δ_{lk} are the cofactors of the determinant. There are N equations corresponding to equation 17 which give the terminal currents. This typical equation may be rewritten into the form

$$I_k = \sum_{l=1}^{N} E_l y_{kl} + I_k' \tag{18}$$

Since the I_k' depends only on the internal generators and the internal elements, these can be measured by shorting all terminals (making the $E_l = 0$) and measuring the kth current. The generalized y parameters may be found from the network which remains when all internal independent generators are set equal to zero (since this makes the I_k' zero).

The generalized Norton form rather than the Thevenin form appeared as a consequence of the loop method applied to the original network. By starting with the node analysis or by simply inverting equations 16 the Thevenin form can easily be found. The special case considered in the text is for $N = 2$.

References

1. SPACE-CHARGE AND TRANSIT-TIME EFFECTS ON SIGNAL AND NOISE IN MICROWAVE TETRODES, L. C. Peterson. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Nov. 1947.

2. TRANSISTOR OPERATION: STABILIZATION OF OPERATING POINTS, R. F. Shea. *Ibid.*, Nov. 1952.

3. TRANSISTOR ELECTRONICS (book), Lo, et al. Prentice-Hall Inc., Englewood Cliffs, N. J., 1955.

4. ANALYSIS AND DESIGN OF TRANSISTOR BIAS NETWORKS, S. K. Ghandi. *Proceedings*, National Electronics Conference, Chicago, Ill., 1956.

A Nonlinear Low-Frequency Instability Phenomenon in Audio Amplifiers

THERON USHER JR.
ASSOCIATE MEMBER AIEE

UNDER certain conditions, audio amplifiers with feedback may exhibit instability at low frequencies, which is not predictable by conventional linear analysis. This instability caused by nonlinear effects can occur when the final stage is in class *B* operation and is preceded by a variable-gain stage, such as a pentode or tetrode.

Mathematical analysis of the problem reveals that a nonminimum phase zero appears in the transfer function of the driver and power stage of the amplifier. This zero, which is not revealed by linear analysis, depends on signal amplitude, and tends to disappear as signal amplitude becomes very small. The presence of the nonminimum phase zero causes an additional 90-degree phase lead at low frequencies in the forward path of the amplifier. This additional phase lead, together with the phase lead from the capacitor-resistor coupling networks between other stages, can therefore cause instability to occur at high-signal amplitudes when feedback is applied to the amplifier.

The remaining sections of the paper contain the theoretical analysis of the problem, experimental results which support the analysis, and a technique for reducing the nonlinear effect.

Analysis

The general assumptions pertaining to the analysis of the problem are listed as follows.

1. Since it has been indicated in the foregoing that the nonlinear effects occur in the variable gain portion of the amplifier and the class *B* power stage, only these two portions will be investigated. For simplicity, the variable-gain portion is treated by considering a single tetrode. The circuit diagram is shown in Fig. 1. The various signal component voltages are defined.

2. The tetrode stage in Fig. 1 is coupled to the power stage by a capacitor-resistor network.

3. The final stage is biased to plate current cutoff.

4. The characteristics of the triode are linear in the conducting regions.

5. The triode power tubes are coupled to a resistive load by an ideal transformer. The winding resistance and leakage inductance are neglected, and the magnetizing inductance is considered infinite.

6. Distortion effects in the tetrode are neglected.

The small-signal transfer function of the configuration in Fig. 1 will first be determined so that the effects of the nonlinearities may easily be seen by comparing the results of the linear analysis with those obtained when the nonlinearities are included.

When the excitation, v_i, consists of a single-frequency sinusoidal voltage, the low-frequency gain of the tetrode is given by equation 1, where g_{mo} is the transconductance at the operating point.

$$\frac{\hat{v}_{1A}}{\hat{v}_i}=-g_{mo}R_1 \triangleq -G_T \tag{1}$$

The coupling network yields the transfer function found in equation 2, where

$$\omega_g=\frac{1}{R_gC}$$

$$\frac{\hat{v}_{2A}}{\hat{v}_{1A}}=\frac{R_gCs}{R_gCs+1}=\frac{s}{s+\omega_g} \quad (R_g>>R_1) \tag{2}$$

The gain of the final stage is given in equation 3, where *N* is the ratio of the turns in one half the primary of the transformer to those in the secondary. The triode parameters are μ and r_p.

$$\frac{\hat{v}_O}{\hat{v}_{2A}}=-\frac{1}{N}\frac{\mu N^2R_L}{r_p+N^2R_L}\triangleq -G_O \tag{3}$$

The over-all transfer function derived by linear analysis is the product of the results in equations 1, 2, and 3.

$$\frac{\hat{v}_O}{\hat{v}_i}=G_OG_T\frac{s}{s+\omega_g} \tag{4}$$

The asymptotic frequency response diagram for the transfer function in equation 4 is found in Fig. 2.

The nonlinear effects of the variable-gain tetrode stage and the class *B* power stage appear when the input signal to the tetrode contains two sinusoidal components. One component has a frequency in the region of interest, near ω_g, and the other component is assumed to have a frequency much higher than ω_g. The input signal is expressed in the following equation.

$$v_i=A_L\sin\omega_Lt+A_H\sin\omega_Ht \tag{5}$$

$$\omega_H>>\omega_g$$

$$A_H>A_L$$

A linear analysis would indicate that the response, v_O, to the input signal in equation 5 can be determined by considering each component separately. However, the following analysis reveals that the response of the amplifier at the frequency

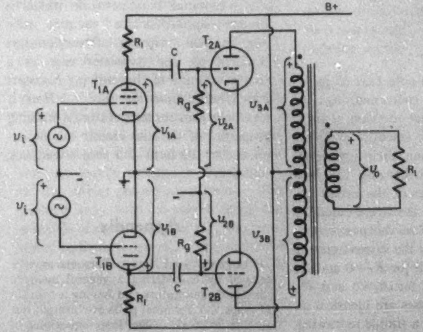


Fig. 1. Driver and power stages of class B amplifier

Vol.
Pt.
195

Paper 57-1029, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted March 7, 1957; made available for printing August 22, 1957.

Theron Usher, Jr., is with Yale University, New Haven, Conn.

The author extends his thanks to the M. B. Manufacturing Company for the use of their facilities, to Galt B. Booth for helpful suggestions, and to Gabriel Hajian for assistance in the experimental work.

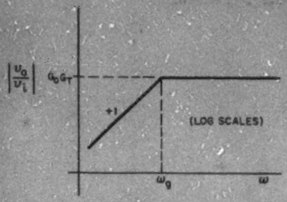


Fig. 2. Asymptotic frequency response, single-frequency input

ω_L is dependent upon A_H, the amplitude of the high-frequency component.

The voltage in equation 5 may be thought of as a high-frequency input to the tetrode, the operating point of which is being varied at the low-frequency rate. Since the transconductance of most tetrodes is nearly linearly dependent upon the operating point, the gain of the tetrode stage to the high-frequency signal component varies at the low-frequency rate. The plate voltage of the tetrode stage contains not only a low-frequency component, but also an amplitude-modulated high-frequency component. The voltage waveform at the plate of tube T_{1A} is shown in Fig. 3. As shown in Fig. 3, the gain of the tetrode stage is highest when the plate voltage is low and vice-versa. The expressions for the signal frequency components of the tetrode plate voltages are given in equations 6 and 6(A).

$$v_{1A} = -\left[g_{mo}R_1A_L \sin \omega_L t + \left(g_{mo} - \frac{dg_m}{dA_L}A_L \sin \omega_L t\right)R_1A_H \sin \omega_H t\right] \quad (6)$$

$$v_{1B} = -v_{1A} \quad (6A)$$

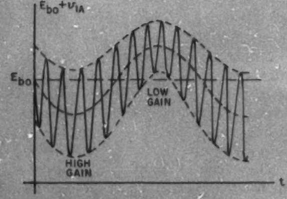


Fig. 3. Tetrode plate voltage waveform

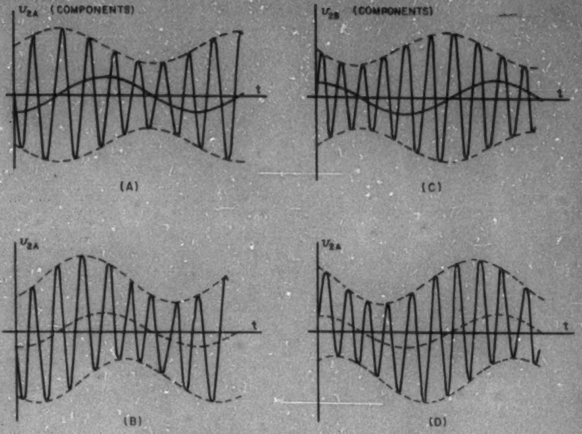


Fig. 4. Grid voltage waveforms, power stage

The upper coupling network in Fig. 1 attenuates the amplitude and shifts the phase of the low-frequency component of v_{1A}, but the modulated high-frequency component is virtually unchanged if $\omega_H >> \omega_g$. The signal frequency components for the grid voltages of tubes T_{2A} and T_{2B} are shown in Fig. 4. Fig. 4(A) depicts the components of v_{2A}. The phase shift of the low-frequency component has been assumed to be 90 degrees. Fig. 4(B) shows the sum of the components of v_{2A}. Note that the upper envelope of v_{2A} is shifted in phase relative to the lower envelope. This is a result of v_{1A} being modified by the coupling network. Figs. 4(C) and (D) show the grid voltage relationships for tube T_{2B}.

Since the power triodes are both biased to plate current cutoff, tube T_{2A} conducts only when v_{2A} is positive, and tube T_{2B} conducts when v_{2B} is positive. Due to the phase inversion from grid to plate, the lower envelope of v_{3A} is the negative of an amplified reproduction of the upper envelope of v_{2A}. Similarly, the lower envelope of v_{3B} is the negative of an amplified reproduction of the upper envelope of v_{2B}. The voltage relationship for the transformer is

$$v_{3A} = -v_{3B} = Nv_O \quad (7)$$

at all times.

From equation 7, the lower envelope of the output voltage, v_O, has the same shape as the negative of the upper envelope of v_{2A}, and the upper envelope of v_O has the same shape as the upper envelope of v_{2B} in Fig. 4. The result for v_O is shown in Fig. 5. It is seen from an inspection of Figs. 5, 4(B), and 4(D) that both envelopes of v_O are in phase, and are equal in magnitude. The output voltage, v_O, therefore consists of a constant-amplitude high-frequency component and a low-frequency component.

A more critical examination of Fig. 5 reveals that the envelopes of v_O are not in phase with the low-frequency component amplified from the grid of T_{2B}. The demodulation effect of the class *B* operated power stage has produced an additional low-frequency voltage component at the plates of T_{3A} and T_{3B}, and also at the transformer secondary. This additional low-frequency component is proportional to the amplified envelope of the modulated high-frequency component at the plate of the tetrode stage.

The equation for the upper envelope of the high-frequency component of v_{1A} is obtained from equations 6 and 6(A) and is found following.

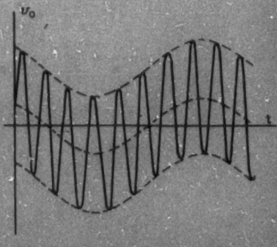


Fig. 5. Output voltage waveforms

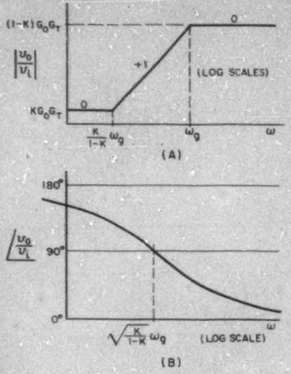


Fig. 6. Amplitude and phase response; 2-frequency input

$$\text{upper envelope of } v_{1A} = +\frac{dg_m}{dA_L} R_L A_L A_H \sin \omega_L t \tag{8}$$

Multiplying equation 8 by $-G_O$, the gain of the final stage, yields the low-frequency component of v_O contributed by the demodulation effect.

$$v_O'' = -G_O \frac{dg_m}{dA_L} R_L A_L A_H \sin \omega_L t \tag{9}$$

The over-all gain for the low-frequency component produced by demodulation is

$$\frac{\hat{v}_O''}{\hat{v}_i} = -G_O \frac{dg_m}{dA_L} R_L A_H \tag{10}$$

The complete expression for the transfer function of the low-frequency signal in the presence of high-frequency signal is found by adding equation 4 and equation 10.

$$\left.\frac{\hat{v}_O}{\hat{v}_i}\right|_L = G_O G_T \frac{s}{s+\omega_g} - G_O \frac{dg_m}{dA_L} R_L A_H \tag{11}$$

The result in equation 11 may be simplified considerably by defining a nondimensional variable, K, in the following fashion

$$K \triangleq \frac{d_{gm}}{dA_L} \frac{A_H}{g_{mo}} < \frac{1}{2} \tag{12}$$

The transfer function for the low-frequency component becomes

$$\left.\frac{\hat{v}_O}{\hat{v}_i}\right|_L = G_O G_T \frac{(1-K)s - K\omega_g}{s+\omega_g} \tag{13}$$

The variable, K, is dependent upon the nonlinear gain characteristics of the tetrode stage, and the amplitude of the high-frequency signal component. By using the nondimensional variable, K, the result in equation 13 is easily extended to include the effect of variable gain occurring in any number of stages preceding the output stage. In general, K is defined

$$K \triangleq \frac{G_T'}{G_T} A_H \tag{12A}$$

where G_T' is the derivative of gain with respect to input signal amplitude, and A_H is the amplitude of the high-frequency signal at the input.

The asymptotic frequency response diagram is found in Fig. 6(A) and the corresponding phase diagram in Fig. 6(B). From Fig. 6(B), it can be seen that the nonminimum phase zero at $(K/1-K)\ \omega_g$ in equation 13 causes an extra 90 degree phase lead at low frequencies. As K increases, the level of the zero-slope asymptote at high frequencies gets lower, and that of the asymptote at low frequencies gets higher until $K = 1/2$, when both asymptotes reach the same level.

If feedback is applied from either the plates of the power triodes, or the secondary of the output transformer, around the gain variable stage and other preceding stages, the nonlinear effects must either be taken into account or effectively eliminated. The feedback amplifier would certainly be stable at low-signal amplitudes, if designed by considering only the linear transfer function in equation 4. However as the signal amplitude is raised, the extra loop gain and phase lead at low frequencies can eventually cause the amplifier to become unstable and exhibit a type of motorboating action in addition to the reproduction of the applied signal.

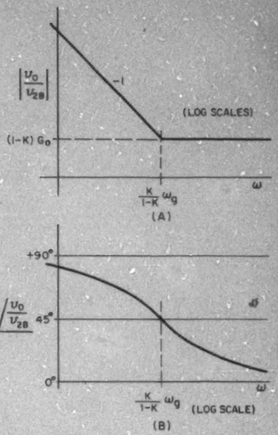


Fig. 8. Amplitude and phase response of power stage

The next section gives a historical note concerning the experimental work performed on a particular amplifier and also the results of experimental tests using signals containing a high-frequency component and a low-frequency component.

Experimental Results

In recent work with high-power audio amplifiers, the baffling type of instability described in the preceding section was noticed. Loop gain tests with single frequency signals revealed that the amplifier should have a wide margin of stability when feedback was applied. However, the motorboating type of instability occurred when the signal amplitude was raised above a certain level, after feedback was applied.

Chronologically, the experimental work with a low-frequency voltage superposed on a high-frequency voltage was performed before the theory in the preceding section was developed. The experimental setup is shown in Fig. 7.

The double frequency input was applied to an early stage of the amplifier. The grid voltage of the upper Eimac 3 × 3000 power triode was applied to

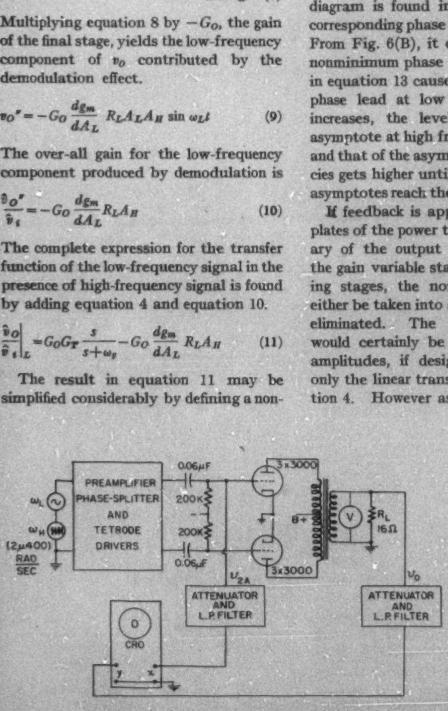

ω_L
ω_H
(2π400) RAD/SEC
PREAMPLIFIER PHASE-SPLITTER AND TETRODE DRIVERS
0.06μF
200K
200K
0.06μF
3x3000
3x3000
B+
V
R_L 16Ω
v_2A
v_O
ATTENUATOR AND L.P. FILTER
ATTENUATOR AND L.P. FILTER
O
CRO
y
x

Fig. 7. Setup for experimental tests

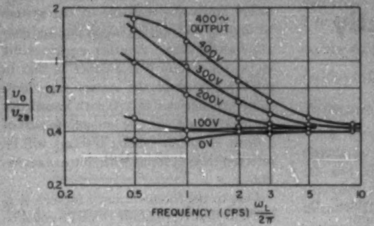
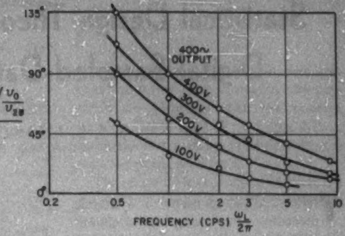


Fig. 9 (left). Experimental amplitude responses for power stage, 2-frequency input

Fig. 10 (right). Experimental phase responses for power stage, 2-frequency input



a low-pass filter and the filter output applied to the horizontal channel of an oscilloscope. The output voltage from the amplifier was applied to an identical low-pass filter, and the filter output applied to the vertical channel of the oscilloscope. The identical low-pass filter sections allowed gain-phase measurements for the final stage to be made directly on the oscilloscope for the low-frequency portion of the signal only.

The theoretical result for the transfer function of the final stage is found by dividing equation 13 by $-G_T\, s/s+\omega_g$.

$$\left.\frac{\hat{v}_O}{\hat{v}_{2B}}\right|_L = -\left.\frac{\hat{v}_O}{\hat{v}_{2A}}\right|_L = G_O\frac{(1-K)s-K\omega_g}{s} \qquad (14)$$

Equation 14 yields the asymptotic gain diagram in Fig. 8(A) and the phase diagram in Fig. 8(B). The experimental results for the frequency response of the low-frequency component for various levels of the 400 cps (cycles per second) high-frequency component are shown in Figs. 9 and 10. From Fig. 10, the phase curves at low frequencies are seen to rise above 90 degrees. It is believed that this is caused by the low-frequency effect of the internal resistance and magnetizing inductance of the output transformer.

The variable K can be calculated from the experimental results by noting the frequency, ω_g', for which there is 45-degree phase shift.

From Fig. 8,

$$\omega_g' = \frac{K}{1-K}\,\omega_g \qquad (15)$$

then

$$K = \frac{\omega_g'}{\omega_g+\omega_g'} \qquad (16)$$

The values of K determined by equation 16 are plotted as a function of the 400 cps output voltage in Fig. 11. The value of K is nearly linearly dependent upon the 400 cps output voltage, as it should be.

Further support of the demodulation theory is afforded by the fact that changing the frequency, but not the amplitude, of the high-frequency component did not affect the gain-phase characteristics of the low-frequency signal. Experimentally, value of the high frequency was varied from 60 to 2,000 cps without changing the results.

Conclusions

It has been shown that the combination of variation in gain with operating point in early amplifier stages and a demodulation phenomenon in the class *B* power stage can cause stability problems when feedback is applied. One solution for combating the stability problem is to reduce the value of K for a given amplitude of high-frequency signal. This essentially means that the variation in tetrode gain with operating point must be reduced.

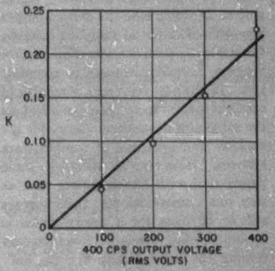


Fig. 11. Nondimensional factor K as a function of the high-frequency output voltage

From the standpoint of practical circuitry, a reduction in gain variation is accomplished by inserting a common cathode resistor in the tetrode stage of Fig. 1. From Fig. 4, the gain of T_{1A} is high when that of T_{1B} is low and vice-versa. The cross-feed caused by the common cathode resistor tends to make the gain per side more nearly uniform, and, furthermore, does not reduce the over-all gain for push-pull operation. The cathode resistor can also be used to provide bias for both tetrodes. This technique may also be used on other variable gain stages in the amplifier.

References

1. THEORY AND APPLICATIONS OF ELECTRON TUBES (book), Herbert J. Reich. McGraw-Hill Book Company, Inc., New York, N. Y., 1944, pp. 223–348.

2. SERVOMECHANISMS AND REGULATING SYSTEM DESIGN (book), H. Chestnut, R. W. Mayer. John Wiley & Sons, Inc., New York, N. Y., 1951.

Statistical Design Theory for Strictly Digital Sampled-Data Systems

S. S. L. CHANG
ASSOCIATE MEMBER AIEE

THIS PAPER presents a synthesis procedure for designing optimum sampled-data systems on a statistical basis. It deals with a strictly digital system, in which case the desired output is undefined except at sampling instants. Consequently such consideration as error between sampling instants is meaningless. In the past few years, a number of excellent papers have been published on the subject of z-transform theory for the analysis and synthesis of sampled-data systems. However, as most of these papers deal with known inputs such as step, ramp, sinusoidal, etc., there appears still to be room for a theory which is applicable when correlation functions concerning the input signals are the only known factors outside the plant to be controlled.

Although the application problems treated in this paper are limited in scope to strictly digital cases without plant saturation, the basic theory and some of the elementary techniques are also useful in formulating a synthesis procedure for digital controlled continuous systems, and for systems with plant saturation and load disturbance. These subjects will be treated in separate papers following the present one.

As expected, the present theory forms a close parallel to Wiener's theory of optimum filtering and prediction, and its various manifestations in the area of continuous feedback control systems. The least-mean-square error criteria, the correlation sequence, the separation of a function into two terms according to root location, etc., can be easily traced to similar concepts and techniques in the continuous case.

Nomenclature

$c(t)$ = controlled variable, or actual system output
$c_1(t)$ = desired output signal
$e_1(t)$ = system error, $C_1(t)-c(t)$
$e_2(t)$ = forcing function, applied directly to plant
$n(t)$ = noise signal, mixed up with reference input
$r(t)$ = reference input
$r_1(t)$ = total input, $r_1(t)=r(t)+n(t)$
$C(z), C_1(z), E_1(z), E_2(z), M(z), R(z), R_1(z)$ = z-transforms of the respective signals
$D(z)$ = transfer function describing the sampled-data processing unit
$F(z)$ = undetermined factor of the system function
$G(z)$ = transfer function describing the plant and hold circuit
$K(z)$ = system function $C(z)/R_1(z)$
$L(z)$ = linear operator relating $c_1(t)$ to $r(t)$
$P(z)$ = product of factors in the numerator of $G(z)$ with zeros outside the unit circle (equation 28)
$Q(z)$ = a meromorphic function of z, to be separated into two terms, $Qi(z)$ and $Q_o(z)$
$Qi(z)$, $Q_o(z)$ = terms in the partial fraction expansion of $Q(z)$ with their poles inside and outside the unit circle respectively (equations 18 and 19)
$Y(z)$ = product of factors of $\Phi_{r_1r_1}(z)$ with poles and zeros inside the unit circle (equation 39)
$\chi(z)$ = an arbitrary function of z with all its poles outside the unit circle
C, C_1, D, etc. = short notations for $C(z)$, $C_1(z)$, $D(z)$, etc. respectively
C^+, C^+_1, D^+, etc. = short notations for $C(z^{-1})$, $C_1(z^{-1})$, $D(z^{-1})$, etc. respectively
m° = number of sampling periods which elapses before the first response of the plant to a step input is noted at a sampling instant
m' = a constant defined in equation 27
q = the probability for a step change not to occur in any one sampling period (Fig. 4)
T = sampling period
$\phi_{11}(n, T)$ = autocorrelation sequence of signal 1
$\phi_{12}(n, T)$ = crosscorrelation sequence of signals 1 and 2
$\Phi_{11}(z)$ = autocorrelation series of signal 1
$\Phi_{12}(z)$ = crosscorrelation series of signals 1 and 2
$\text{ct}\{Q(z)\}$ = constant term in the double ended power series expansion of $Q(z)$
$\underline{\Delta}$ = characteristic determinant, formed by the coefficients on the left hand side of simultaneous equations 96, 97, and 98

The Optimization Problem

Fig. 1 illustrates the general design problem. The controlled variable $c(t)$ is to be as close as possible to $c_1(t)$ which results from a given linear operation on $r(t)$, such as time translation or prediction, integration, or differentiation etc. However, besides the reference input, there are also other undesired inputs to the servo system, e.g., noise $n(t)$ inseparably mixed with $r(t)$. Furthermore, there is to be considered the fixed element or plant, which limits the response of the system because of its transportation lag, and nonminimum phase. Another equivocating factor on system performance is that in some applications the system may have to make a decision on $c(t)$ with knowledge of $r(t)+n(t)$ only in the immediate past of a finite duration. Such systems will be referred to as systems with finite observation.

Due to the forementioned factors, it is obvious that $c(t)$ cannot be made equal to $c_1(t)$ no matter how the compensating elements in the servo system are designed. The mean-square error criteria will be used as a basis for evaluating the relative merits of all possible systems.

For a strictly digital system, it is required that:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N}[c_1(nT)-c(nT)]^2=\text{minimum} \qquad (1)$$

Another basic assumption is that $r(t)$, and $n(t)$ are generated from stationary random processes, so that the ergodic hypothesis is applicable. While this assumption is too restrictive, it may be stipulated that in practice one needs to require only that the change in the generating processes is neglible within a period a few times the system time constants.

Correlation Sequences and Their Transformations

To make a system design on the forementioned basis, the required statistical information about the input signals and disturbances are the correlation sequences. Presently the auto- and crosscorrelation sequences will be defined and their transform properties will be studied. As a useful tool in such studies the double ended z-transform function $A(z)$ of a random signal $a(t)$ is defined as

$$A(z)=\sum_{n=-\infty}^{\infty}a(nT)z^{-n} \qquad (2)$$

Following the established convention,[1] capitalized letters will be used to denote functions in the s and z domains, while lower-case letters denote corresponding time functions. For instance, $A(s)$ denote the Laplace transform of $a(t)$, $A^*(s)$ denote the Laplace transform of the

---

Paper 57-1061, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted May 27, 1957; made available for printing August 19, 1957.

S. S. L. Chang is with New York University, New York, N. Y.

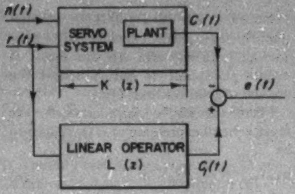


Fig. 1. Diagram showing relation between various variables

sampled function

$$\sum_{n=-\infty}^{n=\infty} a(t)\delta(t-nT),$$

and $A(z)$ denote the z-transform function defined in equation 2. $A(s)$ and $A(z)$ are understood as not being the same functions.

DEFINITION OF CORRELATION SEQUENCES

The autocorrelation sequence $\phi_{11}(n, T)$ of the signal i_1 is defined as

$$\phi_{11}(n, T)=\overline{i_1(mT)i_1(mT+nT)} \tag{3}$$

The foregoing average may be either an ensemble average or a time average over m. Since the signals are assumed to be stationary, there is no difference between the two averages. Similarly, the crosscorrelation sequence between two signals i_1 and i_2 is defined as

$$\phi_{12}(n, T)=\overline{i_1(mT)i_2(mT+nT)} \tag{4}$$

It is obvious that both the correlation sequence and the correlation function can be defined for a continuous signal or signals. They are related as follows:

$$\phi_{11}(n, T)=\phi_{11}(nT) \tag{5}$$

$$\phi_{12}(n, T)=\phi_{12}(nT) \tag{6}$$

SYMMETRICAL PROPERTIES

The following symmetrical properties can be readily established from equations 3 and 4

$$\phi_{11}(n, T)=\phi_{11}(-n, T) \tag{7}$$

$$\phi_{12}(n, T)=\phi_{21}(-n, T) \tag{8}$$

CORRELATION SERIES

The auto and crosscorrelation series are defined as

$$\Phi_{11}(z)=\sum_{n=-\infty}^{n=\infty} \phi_{11}(r, T)z^{-n} \tag{9}$$

$$\Phi_{12}(z)=\sum_{n=-\infty}^{n=\infty} \phi_{12}(n, T)z^{-n} \tag{10}$$

The symmetrical properties of the correlation series follow directly from equations 7 to 10

$$\Phi_{11}(z)=\Phi_{11}(z^{-1}) \tag{11}$$

$$\Phi_{12}(z)=\Phi_{21}(z^{-1}) \tag{12}$$

TRANSMISSION AND SUMMING PROPERTIES

Let signals $a(t)$, and $b(t)$ be the sum signals as shown in Fig. 2, in which H_{ai} (z) and H_{bj} (z) represent the transfer functions of stable systems or prescribed linear operators, and $a_i(t)$ and $b_j(t)$ are stationary random sequences of time, $t=\ldots-2T,\ -T,\ 0,\ T,\ 2T\ldots.$ In terms of z-transform the two signals can be expressed as:

$$A(z)=\sum_i H_{ai}(z)A_i(z) \tag{13}$$

$$B(z)=\sum_j H_{bj}(z)B_j(z) \tag{14}$$

It is shown in Appendix I that:

$$\Phi_{ab}(z)=\sum_{i,j} H_{ai}(z^{-1})H_{bj}(z)\Phi_{a_ib_j}(z) \tag{15}$$

where $\Phi_{ab}(z)$ is the correlation series of the resultant signals $a(t)$ and $b(t)$, and $\Phi_{a_ib_j}(z)$ is that of the source signals $a_i(t)$ and $b_j(t)$.

CLOSED FORM OF A DOUBLE-ENDED POWER SERIES

In mathematical manipulations, it is more convenient to use the closed form of the correlation series whenever such forms are possible. While a single-ended power series expansion is unique for a given rational algebraic expression this is no longer true for double-ended power series expansion, unless some additional restrictions are imposed. The latter can be done as follows: For physical signals without d-c components, the correlation sequences $\phi_{11}(n, T)$ and $\phi_{12}(n, T)$ approach zero as n approaches $\pm\infty$. Thus the coefficients of z^n and z^{-n} at both ends of the series approach zero. This condition makes the double-ended power series expansion unique.

In general, a rational algebraic function $Q(z)$ can be written as a partial fraction expansion in terms of the roots λ_i of the denominator polynomial and a polynomial $\pi(z)$

$$Q(z)=\sum_i \frac{C_i}{z-\lambda_i}+\pi(z) \tag{16}$$

For $|\lambda_i|>1$, $C_i/(z-\lambda_i)$ is expanded in terms of positive powers of z:

$$\frac{C_i}{z-\lambda_i}=-\frac{C_i}{\lambda_i}\sum_{n=0}^{\infty}\left(\frac{z}{\lambda_i}\right)^n$$

For $|\lambda_i|<1$, $C_i/(z-\lambda_i)$ is expanded in terms of negative powers of z:

$$\frac{C_i}{z-\lambda_i}=\frac{C_iz^{-1}}{1-\lambda_iz^{-1}}=C_iz^{-1}\sum_{n=0}^{\infty}\lambda_i^nz^{-n}$$

From the foregoing discussion, one sees that the terms with λ_i outside the unit circle are identified with the positive power series in z, while the terms with λ_i inside the unit circle are identified with the negative power series in z.

The terms with λ_i on the unit circle are treated by a limiting process as shown in example 2.

Summarizing the above, let $Q(z)$ be written alternatively as

$$Q(z)=\sum_{n=-\infty}^{\infty} a_nz^{-n} \tag{17}$$

with the condition that a_n approaches zero as $|n|$ approaches infinity. Let $Q_i(z)$ and $Q_o(z)$ be defined as

$$\left.\begin{aligned} Q_i(z)&\equiv\sum_{n=1}^{\infty} a_nz^{-n}\\ Q_o(z)&\equiv\sum_{n=0}^{\infty} a_{-n}z^{n}\end{aligned}\right\} \tag{18}$$

Then in closed form, $Q_i(z)$ and $Q_o(z)$ are

$$Q_i(z)=\sum_{|\lambda_i|<1}\frac{C_i}{z-\lambda_i},\quad Q_o(z)=\sum_{|\lambda_i|>1}\frac{C_i}{z-\lambda_i}+\pi(z) \tag{19}$$

SIGNIFICANCE OF THE CONSTANT TERM OF $\Phi_{ab}(z)$ AND ITS EVALUATION

The constant term of $\Phi_{ab}(z)$ in its series expansion is simply $\phi_{ab}(O, T)$, which, by definition, is the average value (or expected value) $\overline{a(t)\ b(t)}$. It plays an important role in the optimization procedure, and will be denoted as $\text{ct}\{\Phi_{ab}(z)\}$.

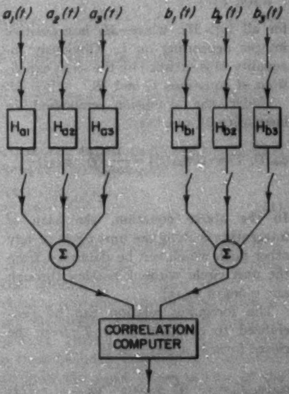


Fig. 2. Diagram of signal transmission

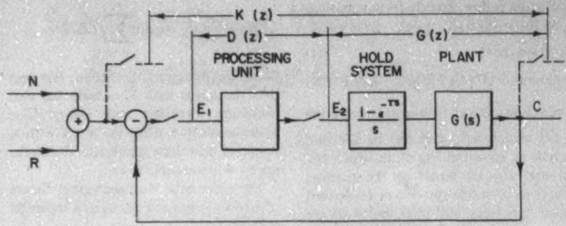


Fig. 3. Block diagram of sampled-data system

When $\Phi_{ab}(z)$ is in closed form, the evaluation of its constant term can be performed by contour integration. Multiplying equation 9 by z^{-1} and integrating with respect to z about a contour enclosing the origin, the series form of $\Phi_{ab}(z)$ yields

$$\sum_{n} \oint \phi_{ab}(n, T)z^{-(n+1)}dz = 2\pi j\phi_{ab}(O, T) \tag{20}$$

The left hand side of equation 20 is equal to

$$\oint [\Phi_{ab}(z)]_{\text{closed}} \frac{dz}{z}$$

if at every point along the path of integration the power series converges uniformly to the closed form $[\Phi_{ab}(z)]$. Because $\phi_{ab}(n, T)\rightarrow 0$ as $|n|\rightarrow \infty$, the series is likely to converge on the unit circle. However, this condition is not sufficient, since there is the possibility of logrithmic divergence. A sufficient condition for uniform convergence on the unit circle is

$$\left|\frac{\phi_{ab}(n+1, T)}{\phi_{ab}(n, T)}\right| \leq l < 1 \tag{21}$$

for all $|n| > M_l$, where M_l is a positive integer depending on l. Obviously inequality 21 is satisfied by the series expansions of equations 17 and 18.

Under the condition embodied in inequality 21, one has

$$\phi_{ab}(O, T) = \text{ct}\{\Phi_{ab}(z)\} = \frac{1}{2\pi j}\oint \Phi_{ab}(z)\frac{dz}{z} \tag{22}$$

In the above equation, the path of integration is along the unit circle or any other path which can be deformed from the unit circle without passing through any poles of $\Phi_{ab}(z)$.

The foregoing relation can be generalized to evaluate $\phi_{ab}(n, T)$ from the closed form of $\Phi_{ab}(z)$. It is

$$\phi_{ab}(n, T) = \frac{1}{2\pi j}\oint_{\text{unit circle}} \Phi_{ab}(z)z^{n-1}dz \tag{23}$$

Basic Optimization Procedure of a Strictly Digital System

Mathematical Formulation

Referring to Fig. 1, $\delta(t)$ is neglected for the time being, $r(t)$ and $n(t)$, are sampled stationary random signals. Since $c(t)$ and $c_1(t)$ are compared only at sampling instants, $c(t)$, $c_1(t)$ and $e(t)$ can also be treated as sampled signals. Let $K(z)$ and $L(z)$ represent respectively the $C(z)/R_1(z)$ ratio of the servo system, and the linear operator which specifies the desired input output relationship. While $L(z)$ is known, $K(z)$ is to be determined such that the mean-square error of equation 1 is to be a minimum. Fig. 1 can be represented mathematically as:

$$\begin{aligned} E(z) &= C_1(z) - C(z) \\ &= L(z)R(z) - K(z)[R(z)+N(z)] \end{aligned} \tag{24}$$

The function $K(z)$ must satisfy the following conditions:

1. $K(z)$ must represent a stable system. Consequently, all its poles are within the unit circle.

2. $K(z)$ must have at least as much transportation lag as the plant. Mathematically, in its power series expansion, $K(z)$ cannot have positive powers of z. Its lowest power of z^{-1} is the same as that of the plant transfer function $G(z)$. Let this factor be denoted as $Z^{-m''}$.

3. $K(z)$ must contain all the zeroes of $G(z)$ which are outside the unit circle. Bergen and Ragazzini derived this condition on the basis that with the slightest error in the cancellation process, the system would be unstable. An alternative view point is that even with mathematically exact cancellation, the system remains inoperative. Let $R_1(z)$ represent the total input $R(z)+N(z)$, and $E_2(z)$ represent the forcing function applied to the plant as shown in Fig. 3. The transfer ratio $E_2(z)/R_1(z)$ is

$$\frac{E_2(z)}{R_1(z)} = \left(\frac{C(z)}{R_1(z)}\right)\left(\frac{E_2(z)}{C(z)}\right) = \frac{K(z)}{G(z)} \tag{25}$$

where

$$K(z) = C(z)/R_1(z)$$

If $K(z)$ does not contain any zero of $G(z)$ which is outside the unit circle, that zero becomes a pole of $E_2(z)/R_1(z)$. Consequently $e_2(t)$ rises exponentially, even though both $r(t)$ and $c(t)$ are zero. It causes the plant to saturate and the system to go into a limit cycle.

From the above considerations, the most general expression of $K(z)$ is seen to be

$$K(z) = z^{-m'}P(z)F(z) \tag{26}$$

where

$$m' = m'' + l \tag{27}$$

$$P(z) = \prod_{i=1}^{i=l} (1 - z/\alpha_i) \tag{28}$$

The l constants α_i's are zeroes of $G(z)$ outside the unit circle. $F(z)$ is the ratio of two polynomials of z^{-1}, and its denominator polynomial must have a nonvanishing constant term. All the roots of $F(z)$ are inside the unit circle. Alternatively one may consider $F(z)$ as a power series in z^{-1}, such that the coefficient of the z^{-i} term approaches zero as i approaches ∞. The two points of view are mathematically equivalent, and will be referred to as condition of physical realizability.

Substituting equation 26 into equation 24, one obtains

$$\begin{aligned} E(z) = [L(z) - z^{-m'}P(z)F(z)]R(z) - \\ z^{-m'}P(z)F(z)N(z) \end{aligned} \tag{29}$$

Equation 1 is equivalent to requiring that $\text{ct}\{\Phi_{ee}(z)\}$ be a minimum. Equations 15 and 29 give

$$\begin{aligned} \Phi_{ee}(z) = [L - z^{-m'}PF] \times \\ [L^+ - z^{m'}P^+F^+]\Phi_{rr} + PP^+FF^+\Phi_{nn} - \\ [L - z^{-m'}PF]z^{m'}P^+F^+\Phi_{nr} - \\ [L^+ - z^{m'}P^+F^+]z^{-m'}PF\Phi_{rn} \end{aligned} \tag{30}$$

The symbols P^+ and F^+ are used to denote $P(z^{-1})$ and $F(z^{-1})$ respectively. The design problem is to determine $F(z)$ such that $\text{ct}\{\Phi_{ee}(z)\}$ is a minimum.

General Solution

Let $F_o(z)$ be the optimum function which gives lowest $\text{ct}\{\Phi_{ee}(z)\}$, $F(z)$ be any other function satisfying the condition of physical realizability, and $F_1(z) = F(z) - F_o(z)$. Obviously $F_1(z)$ also satisfies the condition of physical realizability.

Equation 30 can be written in terms of F_o and F_1

$$\Phi_{ee}(z) = \Phi_{eeo}(z) + \psi_o + \psi_1 + \psi_2 \tag{31}$$

where Φ_{eeo} is the same as the right hand side of equation 30 with F and F^+ replaced by F_o and F_o^+,

$$\psi_o = PP^+F_1F_1^+(\Phi_{rr} + \Phi_{nn} + \Phi_{rn} + \Phi_{nr}) \tag{32}$$

$$\begin{aligned} \psi_1 = (z^{-m'}PF_o - L)z^{m'}P^+F_1^+(\Phi_{rr} + \Phi_{nr}) + \\ PP^+F_oF_1^+(\Phi_{rn} + \Phi_{nn}) \end{aligned} \tag{33}$$

$$\begin{aligned} \psi_2 = (z^{m'}P^+F_o^+ - L^+)z^{-m'}PF_1(\Phi_{rr} + \Phi_{rn}) + \\ PP^+F_o^+F_1(\Phi_{nr} + \Phi_{nn}) \end{aligned} \tag{34}$$

From equation 32, ψ_o can be interpreted as the autocorrelation series of the signal $P^+F_1[R(z)+N(z)]$. Consequently ct $\{\psi_o\}$ is positive definite, unless $F_1=0$. The necessary sufficient condition for $\phi_{ee}(0, T) \geq \phi_{eeo}(O, T)$ with arbitrary F_1 is that

$$\text{ct}\{\psi_1\}+\text{ct}\{\psi_2\}=0 \tag{35}$$

It is readily verified from equations 33 and 34 that $\psi_1(z)=\psi_2(z^{-1})$. Therefore ct $\{\psi_1\}$ = ct $\{\psi_2\}$ and equation 35 can be reduced to

$$\text{ct}\{\psi_1\}=\frac{1}{2\pi j}\oint\frac{\psi_1(z)}{z}\,dz=0 \tag{36}$$

in which the contour integral is taken along the unit circle. Equation 33 can be rewritten as

$$\frac{\psi_1(z)}{z}=F_1^+[PP^+z^{-1}\Phi_{r_1r_1}F_o- Lz^{m'-1}P^+(\Phi_{rr}+\Phi_{nr})] \tag{37}$$

where

$$\Phi_{r_1r_1}=\Phi_{rr}+\Phi_{nn}+\Phi_{rn}+\Phi_{nr}$$

As $F_1(z)$ does not have any pole outside the unit circle, $F_1^+=F_1(z^{-1})$ does not have any pole inside the unit circle. A sufficient condition for satisfying equation 36 is that the expression inside the square bracket of equation 37 does not have any pole inside the unit circle. This is also the necessary condition due to the arbitrariness of $F_1(z)$. If there are poles inside the unit circle, a function F_1 could be selected such that the sum of residues does not vanish, and equation 36 is violated. Therefore

$$PP^+z^{-1}\Phi_{r_1r_1}F_o-LP^+z^{m'-1}(\Phi_{rr}+\Phi_{nr}) =\chi(z) \tag{38}$$

where $\chi(z)$ is an abritrary function with all its poles outside the unit circle. From equation 11, one sees that if μ_i is a pole (or zero) of $\Phi_{r_1r_1}$, $1/\mu_i$ is also a pole (or zero) of $\Phi_{r_1r_1}$. Let $Y(z)$ denote the product of the factors of $\Phi_{r_1r_1}$ with poles or zeros inside the unit circle. Then $Y(z^{-1})$ is the product of the factors of $\Phi_{r_1r_1}$ with poles or zeros outside the unit circle. By properly choosing the constant multiplier, one has

$$Y(z)Y(z^{-1})=\Phi_{r_1r_1}(z) \tag{39}$$

where

$$Y(z)=\frac{A\prod_{i=1}^{M}(1-\beta_iz^{-1})}{\prod_{i=1}^{N}(1-\mu_iz^{-1})} \tag{40}$$

$$Y(z^{-1})=\frac{A\prod_{i=1}^{M}(1-\beta_iz)}{\prod_{i=1}^{N}(1-\mu_iz)}=Y^+ \tag{41}$$

In equations 40 and 41, β_i's are the M zeros of $\Phi_{r_1r_1}(z)$ inside the unit circle and μ_i's are the N poles of $\Phi_{r_1r_1}(z)$ inside the unit circle.

Dividing equation 38 by PY^+, where Y^+ represents $Y(z^{-1})$, it is

$$z^{-1}P^+YF_o-LP^+z^{m-1}(\Phi_{rr}+\Phi_{nr})/PY^+ =\chi(z)/PY^+ \tag{42}$$

Let $Q(z)$ represent the expression:

$$Q(z)=LP^+z^{m'-1}(\Phi_{rr}+\Phi_{nr})/PY^+ \tag{43}$$

Separating $Q(z)$ into two parts $Q_i(z)$ and $Q_o(z)$ as shown by equations 18 and 19, equation 42 becomes

$$z^{-1}P^+YF_o-Q_i(z)=Q_o(z)+\chi(z)/PY^+ \tag{44}$$

All the poles on the left hand side of equation 44 are inside the unit circle while all the poles on the right hand side of equation 44 are outside the unit circle. In order to satisfy equation 44 both sides must equal to a constant C. Therefore:

$$z^{-1}P^+YF_o-Q_i(z)=C \tag{45}$$

Rearranging equation 45, one obtains

$$F_o=\frac{zQ_i(z)}{P^+Y}+\frac{Cz}{P^+Y} \tag{46}$$

From equation 18, $Q_i(z)$ does not have a constant term, and $F_o(z)$ does not have a z term. Consequently, C must be zero. The final solution is

$$F_o(z)=\frac{zQ_i(z)}{P(^{-1})Y(z)} \tag{47}$$

$$K(z)=\frac{z^{-m'+1}P(z)Q_i(z)}{P(z^{-1})Y(z)} \tag{48}$$

Summarizing the above, from the correlation series of the reference input and noise and the plant transfer function, $Y(z)$ and $Q_i(z)$ are calculated from equations 40 and 43. The closed-loop system function $K(z)$ is then determined according to equation 48. Once $K(z)$ is known, $D(z)$ and the processing unit can be readily synthesized.

Example 1: Optimum Filtering and Prediction

Referring to Fig. 3, let

$$G(s)=\frac{10}{1+2Ts} \tag{49}$$

$$L(z)=z^n \tag{50}$$

$$\Phi_{nn}(z)=1 \tag{51}$$

$$\Phi_{rr}(z)=\frac{A^2(1-g^2)}{(1-gz^{-1})(1-gz)}\quad g<1 \tag{52}$$

Equation 50 represents desired prediction of n-sampling periods ahead. Equation 51 represents white Gaussian noise, and equation 52 represents random-step input signal as shown in Fig. 4. In Fig. 4 the average value of $[r(t)^2]$ is A^2; the amplitudes before and after each change are uncorrelated, and the probability of change in each sampling period is $1-q$. The derivation of equation 52 from Fig. 4 is given in Appendix II. Inputs $r(t)$ and $n(t)$ are assumed to be uncorrelated, or $\Phi_{rn}(z)=0$.

From equation 15

$$\Phi_{r_1r_1}=\Phi_{rr}+\Phi_{nn}=\frac{g}{a}\frac{(1-az^{-1})(1-az)}{(1-gz^{-1})(1-gz)} \tag{53}$$

where a is a constant given by:

$$a+\frac{1}{a}=\frac{A^2(1-g^2)}{g}+g+\frac{1}{g},\quad |a|<1 \tag{54}$$

Therefore

$$Y=\sqrt{\frac{g}{a}}\,\frac{1-az^{-1}}{1-gz^{-1}} \tag{55}$$

From equation 49 and Fig. 3, the system function describing the hold circuit and plant is

$$G(z)=\frac{3.94z^{-1}}{1-0.606z^{-1}} \tag{56}$$

Therefore $P(z)=1$, and $m''=m'=1$. Equation 43 gives

$$Q(z)=z^n\Phi_{rr}/Y^+=\frac{z^{n+1}A^2(1-g^2)\sqrt{a/g}}{(z-g)(1-az)} \tag{57}$$

As both a and q are less than unity in magnitude, the only term in the partial fraction expansion which has a pole inside the unit circle is the $z-q$ term. Consequently

$$Q_i(z)=\frac{A^2g^{n+1}(1-g^2)\sqrt{a/g}}{(z-g)(1-ag)} \tag{58}$$

Equation 48 gives:

$$K(z)=\frac{A^2ag^n(1-g^2)z^{-1}}{(1-az^{-1})(1-ag)} \tag{59}$$

Equation 54 can be written as

$$\frac{A^2(1-g^2)}{g}=a+\frac{1}{a}-g-\frac{1}{g}=(g-a)(1-ag)/ag \tag{60}$$

Combining equations 59 and 60, the final form of $K(z)$ is obtained

$$K(z)=\frac{(g-a)g^nz^{-1}}{1-az^{-1}} \tag{61}$$

The physical meaning of equation 61 is as follows: The factor $1/(1-az^{-1})$ gives the filtering or averaging effect. The closer a is to unity, the more digits are being effectively included in the averaging process. However, as a is always less than q, these included sampling digits are likely to have the same value as that of the present. Fig. 5 illustrates the filtering constant a as a function of the signal to

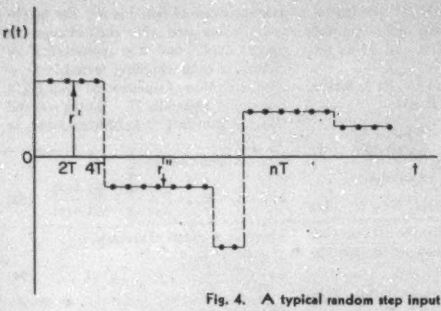


Fig. 4. A typical random step input

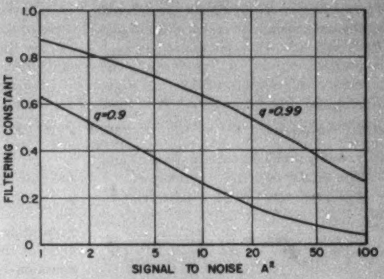


Fig. 5. Noise-filtering constant versus signal to noise ratio

noise ratio A^2 for two different values of q: 0.9 and 0.99. It is noted that a reduces as A^2 is increased.

From equation 61

$$D(z)G(z)=\frac{K}{1-K}=\frac{(q-a)q^n z^{-1}}{1-[a+(q-a)q^n]z^{-1}} \tag{62}$$

The digital processing unit is given as

$$D(z)=\frac{(q-a)q^n}{3.94}\cdot\frac{1-0.606z^{-1}}{1-[a+(q-a)q^n]z^{-1}} \tag{63}$$

EXAMPLE 2: LIMITING PROCESSES

As a second example, still referring to Fig. 3, let $T=1$, and

$$G(s)=\frac{10}{s(1+s)^2} \tag{64}$$

$$L(z)=z^n \tag{65}$$

$$\Phi_{nn}(z)=0 \tag{66}$$

Let $r(t)$ be a random series of ramp inputs obtained from integrating $r(t)$ of Fig. 4, as shown in Fig. 6. To avoid confusion, $r(t)$ of Fig. 4 is denoted as $r_4(t)$. To overcome the difficulties of having a root on the unit circle, a small constant γ is introduced. One may visualize γ as the reciprocal of one million years. For all practical purposes, Fig. 6 is the same as straight integration, while mathematically there is not any root on the unit circle. Thus our limiting process is physically justified. From equation 15

$$\Phi_{rr}=\frac{B^2}{(1-gz)(1-gz^{-1})(1-bz)(1-bz^{-1})} \tag{67}$$

where

$B^2=A^2(1-q^2)$, and $b=e^{-\gamma T}$

Bergen and Ragazzini[2] calculated $G(z)$ to be

$$G(z)=\frac{z^{-1}(1+2.34z^{-1})(1+0.16z^{-1})}{(1-z^{-1})(1-0.368z^{-1})^2} \tag{68}$$

Hence $m'=2$, and $P(z)=1+z/2.34=1+0.427z$. Since $\Phi_{nn}(z)=0$, $\Phi_{r_1r_1}(z)=\Phi_{rr}(z)$ and

$$Y(z)=\frac{B}{(1-gz^{-1})(1-bz^{-1})} \tag{69}$$

Equation 43 gives:

$$Q(z)=\frac{Bz^{n+1}(1+0.427z^{-1})}{(1+0.427z)(1-gz^{-1})(1-bz^{-1})}$$
$$=\frac{BU(z)}{(z-g)(z-b)} \tag{70}$$

where $U(z)=z^{n+2}\,(z+0.427)/(1+0.427z)$. By partial fraction expansion,

$$Q_1(z)=\frac{B[U(b)(z-g)-U(g)(z-b)]}{(b-g)(z-g)(z-b)}$$
$$=\frac{Bz^{-1}(\alpha-\beta z^{-1})}{(1-gz^{-1})(1-bz^{-1})} \tag{71}$$

where

$$\alpha=\frac{U(b)-U(g)}{b-q} \tag{72}$$

$$\beta=\frac{qU(b)-bU(g)}{b-g} \tag{73}$$

From equation 48 the optimum system function is determined to be

$$K(z)=\frac{z^{-1}[0.427+z^{-1}][\alpha-\beta_z{}^{-1}]}{1+0.427z^{-1}} \tag{74}$$

By letting γ approach zero, b approaches unity, and equations 72 and 73 become

$$\alpha=\frac{1-U(g)}{1-q} \tag{75}$$

$$\beta=\frac{q-U(g)}{1-q}=\alpha-1 \tag{76}$$

Equations 74, 75, and 76 give the optimum system function with the frequency of changing slope taken into account. It can be readily verified from the definition of q that the average time for $r(t)$ to have the same slope is $1/(1-q)$ sampling periods. Corresponding to the case of a ramp input of infinite duration, q approaches unity, and

$$\alpha=\left.\frac{dU(g)}{dq}\right|_{q=1}=\frac{1}{U(q)}\left.\frac{dU(q)}{dq}\right|_{q=1}$$
$$=\left[\frac{n+2}{q}+\frac{1}{q+0.427}-\frac{0.427}{1+0.427q}\right]_{q=1}$$
$$=n+2+\frac{1-\lambda}{1+\lambda} \tag{77}$$

where $\lambda=0.427$ in this special case. Equation 74 can be written as

$$K(z)=\frac{z^{-1}(\lambda+z^{-1})(3-\lambda-2z^{-1})}{(1+\lambda z^{-1})(1+\lambda)}+\frac{nz^{-1}(\lambda+z^{-1})(1-z^{-1})}{1+\lambda z^{-1}} \tag{78}$$

The first term on the right hand side of equation 78 represents an optimum system function without prediction, and will be denoted as $K_o(z)$. The second term represents an expected addition which brings $c(t)$ to that of n-sampling periods later. To verify that the system error is eventually zero, $1-K_o(z)$ is found to be

$$1-K_o(z)=\frac{(1-z^{-1})^2[1+2z^{-1}/(1+\lambda)]}{1+\lambda z^{-1}} \tag{79}$$

From equations 78 and 79, for $n=0$,

$$D(z)G(z)=\frac{K_o(z)}{1-K_o(z)}$$
$$=\frac{z^{-1}(\lambda+z^{-1})(3-\lambda-2z^{-1})}{(1-z^{-1})^2(1+\lambda+2z^{-1})} \tag{80}$$

The digital processing unit is specified by

$$D(z)=\frac{0.771(1+0.777z^{-1})(1-0.368z^{-1})^2}{(1-z^{-1})(1+1.40z^{-1})(1+0.16z^{-1})} \tag{81}$$

If a unit ramp is applied to the system,

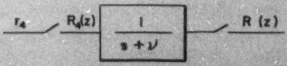


Fig. 6. Derivation of ramp input from step input

the system error is obtained from equation 79 as

$$E_1(z)=[1-K_0(z)]R(z)$$

$$=\frac{z^{-1}[1+2z^{-1}/(1+\lambda)]}{1+\lambda z^{-1}} \quad (82)$$

The total sum of the squares of system error at each sampling instant is

$$\sum_n [e_1(n, T)]^2=\frac{1}{2\pi j}\oint E_1(z)E_1(1/z)\frac{dz}{z}$$

$$=\frac{5+3\lambda}{(1+\lambda)^2} \quad (83)$$

For the present problem, this value is 2.16. It is the lowest obtainable, no matter what $D(z)$ is used.

Optimization of a System With Finite Observations, General Procedure

THEORY

A system limited to n' previous observations can be described by the system function

$$K(z)=\sum_{i=m''}^{n'} a_i z^{-i} \quad (84)$$

The lower limit m'' is due to the plant lag, and the upper limit n' is due to the limit on the number of previous observations. As before,

$$K(z)=z^{-m'}P(z)F(z) \quad (85)$$

The constant m' and $P(z)$ are defined the same way as in the previous section. However, due to equation 84,

$$F(z)=\sum_{i=0}^{i=n'-m'} b_i z^{-i} \quad (86)$$

Following the same steps as before, one arrives at an equation identical with equation 36

$$c_1t_1\{\psi_1\}=ct\{F_1^+[PP^+\Phi_{r_1r_1}F_0-Lz^{m'}P^+(\Phi_{rr}+\Phi_{nr})]\}$$

$$=ct\{F_1^+z\chi(z)\}=0 \quad (87)$$

Since F_1^+ may have arbitrary coefficients of z^i, $i=0, 1, 2, \ldots n'-m'$, equation 87 is mathematically equivalent to requiring that all the coefficients of z^{-i}, $i=0, 1, 2 \ldots n'-m'$, in the double-ended power-series expansion of $z\chi(z)$ vanish. Let the known functions in $\chi(z)$ be expanded into series form as follows:

$$PP^+\Phi_{r_1r_1}=\sum_{i=-\infty}^{\infty} \xi_i z^{-i} \quad (88)$$

$$Lz^{m'}P^+(\Phi_{rr}+\Phi_{nr})=\sum_{i=-\infty}^{\infty} \eta_i z^{-i} \quad (89)$$

Equation 87 is equivalent to requiring the coefficients b_i's to satisfy the following:

$$\sum_{i=0}^{i=n'-m'}\sum_{j=0}^{n'-m'} b_i\xi_{j-i}z^{-j}-\sum_{j=0}^{n-m'}\eta_j z^{-j}=0 \quad (90)$$

Equating coefficients of z in equation 90,

$$\sum_{i=0}^{i=n'-m'} b_i\xi_{j-i}=\eta_j,\ j=0, 1, 2\ldots n'-m' \quad (91)$$

The above $n'-m'+1$ equations allow the coefficients b_i's to be uniquely determined, provided that the determinant $\underline{\Delta}=|\xi_{j-i}|\neq 0$. Due to the definition of $\Phi_{r_1r_1}$, P, and P^+, $\xi_{j-i}=\xi_{i-j}$. Furthermore, only the first $n'-m'+1$ coefficients ξ's and η's need to be evaluated. The physical interpretation is simply that since the system is not allowed to utilize more than $n'-m'+1$ input values simultaneously, statistical knowledge beyond that is unnecessary.

EXAMPLE 3

In Example 1, the number of past observations, n', is unlimited. Suppose that $n'=3$, what is then the optimum system function?

Writing $\Phi_{rr}(z)$ into series form,

$$\Phi_{rr}(z)=A^2\sum_{i=-\infty}^{\infty} g^i z^{-i} \quad (92)$$

Since $\Phi_{nn}(z)=1$, the values of the ξ's are

$$\xi_o=1+A^2$$

$$\xi_i=\xi_{-i}=A^2g^i \quad i=1, 2, 3\ldots \quad (93)$$

Let $L(z)=z^n$, and equation 89 gives

$$A^2z^{n+1}\sum_{i=-\infty}^{\infty} g^i z^{-i}=\sum_{n=-\infty}^{\infty}\eta_i z^{-i} \quad (94)$$

Equating coefficients of z^{-i}, one obtains

$$\eta_i=A^2q^{i+n+1}, \quad i=0, 1, 2, 3\ldots \quad (95)$$

Equation 91 can be written as:

$$\left(1+\frac{1}{A^2}\right)b_o+qb_1+q^2b_2=q^{n+1} \quad (96)$$

$$qb_o+\left(1+\frac{1}{A^2}\right)b_1+qb_2=q^{n+2} \quad (97)$$

$$q^2b_o+qb_1+\left(1+\frac{1}{A^2}\right)b_2=q^{n+3} \quad (98)$$

The solution of the above equations is

$$b_0=q^{n+1}-\frac{q^{n+1}}{A^2\underline{\Delta}}\left[\left(1+\frac{1}{A^2}\right)^2-q^2\right] \quad (99)$$

$$b_1=\frac{q^{n+2}}{A^2\underline{\Delta}}\left[1+\frac{1}{A^2}-q^2\right] \quad (100)$$

$$b_2=\frac{q^{n+3}}{A^4\underline{\Delta}} \quad (101)$$

where _ is the determinant of the coefficients in equations 96, 97 and 98. The optimum closed loop system function is

$$K(z)=b_0z^{-1}+b_1z^{-2}+b_2z^{-3} \quad (102)$$

Conclusion

The present design theory of sampled-data systems are based on system response to inputs of known shape. Little attention has been given to situations in which only the statistical properties of the inputs are known. To fill this void somewhat, this paper presents the basic ingredients of a statistical design theory, and applies it to the optimum design of strictly digital systems. More specifically, the following topics have been treated:

1. Definitions of correlation sequences and correlation series for sampled signals.
2. The transmission, and summing properties of correlation series.
3. Evaluation of the correlation sequence from the closed form of the correlation series.
4. Optimization of a strictly digital system based on the mean-square error criteria.
5. The same optimization problem with the additional constraint of finite observation.
6. Illustrative examples of optimum prediction in the presence of noise, and of a limiting process.

Appendix I. Proof of the Transmission and Summing Properties of Correlation Series

Let $A_m(z)$ and $B_m(z)$ represent the z-transforms of the sampled signals $a(t)$ and $b(t)$ respectively, where $t=0, T, 2T, \ldots mT$. Then

$$A_m(z^{-1})B_m(z)=\sum_{i=0}^{m} a(iT)z^i\sum_{j=0}^{m} b(jT)z^{-j}$$

$$=\sum_{i=0}^{m}\sum_{k=-i}^{k=m-i} a(iT)b(iT+kT)z^{-k}$$

$$=\left(\sum_{k=-m}^{-1}\sum_{i=|k|}^{m}+\sum_{k=0}^{m}\sum_{i=0}^{m-k}\right)\times a(iT)b(iT+kT)z^{-k} \quad (103)$$

Dividing equation 103 by m and taking ensemble average, there results

$$\frac{1}{m}\overline{A_m(z^{-1})B_m(z)}=\sum_{k=-m}^{m}\left(1-\frac{|k|}{m}\right)+\phi_{ab}(k, T)z^{-k} \quad (104)$$

Letting m approach infinity, equation 104 becomes:

$$\lim_{m\to\infty}\frac{1}{m}\overline{A_m(z^{-1})B_m(z)}=\sum_{k=-\infty}^{\infty}\phi_{ab}(k_1T)^{-k}$$

$$=\Phi_{ab}(z) \quad (105)$$

Referring to Fig. 2, let $A_m'(z)$ and $B_m'(z)$ represent the z-transforms of the sum signals $a'(t)$ and $b_m'(t)$ when the source signals are limited to m sampled sets of values. Obviously $a(t)$ and $b(t)$ persist somewhat longer due to the storage properties of $H_{ai}(z)$ and $H_{bj}(z)$. However this end effect has only a limited number of significant digits, and is negligible if m is made large enough. From equations 13 and 14, one obtains

$$A_m'(z^{-1})B_m'(z)=\sum_i\sum_j\times H_{ai}(z^{-1})H_{bj}(z)A_{mi}(z^{-1})B_{mj}(z) \quad (106)$$

Dividing equation 106 by m, taking ensemble average and letting m approach infinity, there results

$$\lim_{m\to\infty}\frac{1}{m}\overline{A_m'(z^{-1})B_m'(z)}=\sum_i\sum_j\times H_{ai}(z^{-1})H_{bj}(z)\Phi_{aibj}(z) \quad (107)$$

Since the end effects become negligible as m approaches infinity, one has

$$\lim_{m\to\infty}\frac{1}{m}\overline{A_m'(z^{-1})B_m'(z)}=\lim_{m\to\infty}\frac{1}{m}\overline{A_m(z^{-1})B_m(z)}=\Phi_{ab}(z) \quad (108)$$

Equation 15 follows from equations 107 and 108.

Appendix II. Calculation of Correlation Series $\phi(z)$ from Random Step Inputs

The average (or expectation) value of $r(nT)r(nT+mT)$ is

$$\phi_{rr}(m,T)=q^m\overline{[r(nT)]^2}=q^mA^2 \quad (109)$$

In the above equation, q^m is the probability that no change has occurred in the m-sampling periods. Referring to Fig. 4, since the correlation between two different amplitudes r' and r'' is assumed to be zero, the only possible event which contributes to $\phi_{rr}(m, T)$ is that no change has occurred.

By definition of $\Phi_{rr}(z)$, it is then

$$\Phi_{rr}(z)=A^2\sum_{i=-\infty}^{\infty}q^{|i|}z^{-i}$$

$$\Phi_{rr}(z)=A^2\sum_{i=1}^{\infty}q^iz^{-i}+A^2\sum_{i=0}^{\infty}q^iz^i$$

$$=\frac{A^2qz^{-1}}{1-qz^{-1}}+\frac{A^2}{1-qz}$$

$$=\frac{A^2(1-q^2)}{(1-qz^{-1})(1-qz)} \quad (110)$$

Equation 110 is the same as equation 52.

References

1. Control System Synthesis (book), **John G. Truxal.** McGraw-Hill Book Company, Inc., New York, N. Y., 1955, pp. 410–558.

2. Sampled-Data Processing Techniques for Feedback Control Systems, **Arthur R. Bergen, John R. Ragazzini.** *AIEE Transactions*, vol. 73, pt. II, Nov. 1954, pp. 236–47.

3. A New Approach to the Design of Pulse Monitored Servo Systems, **A. Porter, F. Stoneman.** *Journal*, Institution of Electrical Engineers, London, England, vol. 97, pt. II, 1950, pp. 597–610.

4. Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation, **William K. Linvill.** *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1779–88.

5. The Analysis of Sampled-Data Systems, **J. R. Ragazzini, L. A. Zadeh.** *Ibid.*, vol. 71, pt. II, Nov. 1952, pp. 225–34.

6. A General Theory of Sampling Servo-Systems, **D. F. Lawden.** *Proceedings*, Institution of Electrical Engineers, vol. 98, pt. IV, Oct. 1951, pp. 31–36.

7. Analysis and Synthesis of Sampled-Data Control Systems, **Eliahu I. Jury.** *AIEE Transactions*, vol. 73, pt. I, Sept. 1954, pp. 332–46.

8. The Pulse Transfer Function and Its Application to Sampling Servo Systems, **R. H. Barker.** *Proceedings*, Institution of Electrical Engineers, vol. 99, pt. IV, Dec. 1952, pp. 302–17.

9. Analysis of Control Systems Involving Digital Computers, **W. K. Linvill, J. M. Salzer.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, no. 7, July 1953, pp. 901–06.

10. The Design of Sampled-Data Feedback Systems, **Gladwyn V. Lago, John G. Truxal.** *AIEE Transactions*, vol. 73, pt. II, Nov. 1954, pp. 247–53.

11. Extension of Continuous-Data System Design Techniques to Sampled-Data Control Systems, **G. W. Johnson, D. P. Lindorff, C. G. A. Nordling.** *Ibid.*, vol. 74, pt. II, Sept. 1955, pp. 252–63.

12. Additions of z-Transformation Theory for Sampled-Data Systems, **Gladwyn V. Lago.** *Ibid.*, vol. 73, pt. II, 1954 (Jan. 1955 section), pp. 403–08.

13. Analysis of Errors in Sampled-Data Feedback Systems, **Jack Sklansky, J. R. Ragazzini.** *Ibid.*, vol. 74, pt. II, May 1955, pp. 65–71.

14. Factors in the Design of Digital Controllers for Sampled-Data Feedback Systems, **John E. Bertram.** *Ibid.*, vol. 75, pt. II, July 1956, pp. 151–59.

15. The Effect of Pole and Zero Locations on the Transient Response of Sampled-Data Systems, **Eliahu I. Jury.** *Ibid.*, vol. 74, pt. II, Mar. 1955, pp. 41–48.

16. Correlation Between Root-Locus and Transient Response of Sampled-Data Control Systems, **Eliahu I. Jury.** *Ibid.* (Jan. 1956 section), pp. 427–35.

17. Synthesis and Critical Study of Sampled-Data Control Systems, **Eliahu I. Jury.** *Ibid.*, vol. 75, pt. II, July 1956, pp. 141–51.

18. The Extrapolation, Interpolation, and Smoothing of Stationary Time Seriel (book), **N. Wiener.** John Wiley & Sons, Inc., New York, N. Y., 1948.

19. An Extension of Wiener's Theory of Prediction, **L. A. Zadeh, J. R. Ragazzini.** *Journal of Applied Physics*, New York, N. Y., vol. 21, July 1950, pp. 645–55.

20. Compensation of Feedback Control Systems Subject to Saturation, **G. C. Newton, Jr.** *Journal*, Franklin Institute, Philadelphia, Pa., vol. 254, Oct. 1952, pp. 281–96; Nov. 1952, pp. 391–413.

21. Linear Filtering of Sampled Data, **Gene Franklin.** *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 4, 1955, pp. 119–28.

◆

Discussion

Arthur R. Bergen (Columbia University, New York, N.Y.): This paper is a comprehensive treatment of a number of interesting synthesis problems in the area of strictly digital systems. It treats with both the finite and infinite memory versions of optimum filter and optimum control system synthesis. Of particular interest is the synthesis of an optimum sampled-data control system when the plant has an associated transportation lag or when the pulsed transfer function of the plant has zeros outside of the unit circle. The author shows that in these cases an optimum $K(z)$ may be synthesized subject to suitable constraints. This is a worthwhile contribution to the sampled-data literature.

It is felt, however, that the utility of the paper would have been enhanced had the close parallel to Wiener's theory, alluded to by the author, been exploited more consistently in the presentation of the material. From the readers' point of view this would have the important advantage of relating a new and unfamiliar problem to one already solved, and concomitantly, ease considerably the burden of notation.

For example, the problem of synthesizing $K(z)$ subject to constraints imposed by a transportation lag or zeros of $G(z)$ outside the unit circle may be easily reduced to a conventional Wiener problem. This may be seen by the manipulation of Fig. 1.

The constrained portion of $K(z)$ [in the author's notation this is $z^{-m'}P(z)$] and its reciprocal may be cascaded with $L(z)$ without changing the system. But now $z^{-m'}P(z)$ is common to the $K(z)$ and $L(z)$ channels and may equivalently be moved to the input signal lines. The problem is thus reduced to one of design of an optimum filter $F_0(z)$ (subject only to the constraint of physical realizability) with respect to a new desired linear operator represented by $L(z)$ in cascade with the reciprocal of $z^{-m'}P(z)$ and with a new pair of inputs simply related to the original ones.

It should be noted that the success in designing $K(z)$ with a prescribed factor $P(z)$ depends on the location of the zeros of $P(z)$ outside the unit circle. If, for some reason, an attempt were made to incorporate a prescribed factor having zeros (or poles) inside the unit circle, this factor would be cancelled in the course of the optimization procedure. No cancellation problem is encountered, however, in dealing with a transportation lag.

Frederick W. Nesline, Jr. (Massachusetts Institute of Technology, Cambridge, Mass.): Professor Chang has indicated the close parallel that exists between optimum design theory in strictly digital systems and in strictly continuous systems. I wish to emphasize this parallel and point out that all of Professor Newton's work[1] on optimization of linear, continuous systems is directly transferable to the digital case. In particular, either an integral-squared error or a mean-squared-error criterion can be used, depending upon whether the input is a transient signal or a stochastic signal. Also, if the technique of minimization of one variable in the presence of a constraint on another variable is used, optimum design can be accomplished in cases where the system has both types of inputs. In addition, it is possible to show that one must accept definite performance limitations if the fixed system includes saturation or elements that have zeros (finite or otherwise) outside the unit circle.

It should be noted that the optimum linear theory discussed is directly dependent upon the input signal's characteristics and the fixed system's characteristics. If either of these items change, the optimum compensator changes. An example of this effect (using a different definition for optimum) was presented by the discusser at the October 1957 Computers in Control Conference at Atlantic City, N. J.

Since the characteristics of physical systems and their input signals often depend upon environment, a true optimum system must be self-adaptive. The application of sampled data theory to this case is promising, but since such a system must always be nonlinear, no completely general solution is possible. An example of one type of optimum self-adaptive system is now being investigated as thesis under the direction of the discusser. This discussion is presented to indicate that optimum theory is not exhausted and that considerable effort in this direction is still necessary before an optimum self-adaptive theory can be specified.

REFERENCE

1. Analytical Design of Linear Feedback Controls (book), G. C. Newton, Jr., L. A. Gould, J. F. Kaiser. John Wiley & Sons, Inc., New York, N. Y., 1957.

S. S. L. Chang: I wish to thank Professors Bergen and Nesline for their interesting and constructive comments.

There is a physical interpretation of Professor Bergen's suggestion of cascading the reciprocal of $s^{-m} P(s)$ with $L(s)$: The problem then becomes prediction or approximating the response of an unstable system. Since a stable physical system can never truly predict nor truly approximate the response of an unstable system, some interesting compromise must be made. His remarks about the impossibility of attaching an arbitrary zero inside the unit circle to $P(s)$ can be interpreted in the same light. $F_0(s)$ is then required to approximate a pole inside the unit circle. Since a stable system can be approximated as closely as desired, the limit of closeness is always reached by having exactly the same pole in $F_0(s)$.

Professor Nesline commented about the performance limitations due to zeroes outside the unit circle. This is already taken care of by the factor $P(s)$, as far as the optimum system function $K_0(s)$ is concerned. The least-mean-square error due to this effect can be readily calculated from equation 30 once $K_0(s)$ is known. While Newton's work of introducing constraint was not expressively assimulated in the present paper, it should not be beyond anyone who has read Newton's paper (reference 19 of the paper). An illustration of this point is given in the second paper of the series.[1]

REFERENCE

1. Statistical Design Theory for Digital-Controlled Continuous Systems, S. S. L. Chang. *AIEE Transactions*, vol. 77, pt. II, 1958 (*Paper no. 58-90*).

Russian to English Machine Translation With Simple Logical Processing

ROBERT E. WALL, JR.
ASSOCIATE MEMBER AIEE

UDO K. NIEHAUS
NONMEMBER AIEE

A SKILLED polyglot accomplishes translation from one language to another by translating thoughts; the source-language material is read and digested, and then the thought of the material is formed into target language sentences and recorded. The unskilled translator would translate first on a word-for-word basis, and would then make improvements in the output by consideration of the context of the words. The ultimate in machine translation would be that of translation by thought, as the accomplished polygot would translate, but several fundamental problems remain to be solved before anything of this sort may be accomplished. For instance, there is a great spread between the amount of storage in the human brain and the greatest amount of storage available in any present-day computer. The brain is estimated to have a total storage capacity of 10^{15} to 10^{20} bits. The largest of computers stores only about 10^9 bits, or not over one millionth as much storage as the human brain. A second limitation is that of present knowledge of language. Linguistic knowledge concerning many problems which are vital to machine translation is rather lacking. Thus it would seem that in the present elementary state of the machine translation art the attempt to perform translations should be made more as the unskilled translator does, by basing the translations on the word-for-word translation, and making such improvements as possible from a consideration of the context of the words in the sentence.

There are three general ways in which word-for-word translations may be improved by logical processing of the material:

1. By grammatical considerations of the usages of the words; first, in order that proper English prepositions may be inserted into the output, in and only in the proper places; and second, to solve problems of number (whether a word is singular or plural). The problem of placement of English prepositions is especially important in translating from highly inflected languages such as Russian, Greek, or Latin, where the functions of many English prepositions are performed by an inflection of the ending of the particular substantive. Problems of number enter into the translation effort because some endings may denote either singular or plural, depending on the grammatical usage of the word in a sentence. The problem of determining the grammatical usage is called that of grammatical incident meaning.

2. By consideration of the meaning or thought of a particular sentence in order that those source-language words which have more than one English equivalent may have the proper equivalent selected. For instance in the case of the Russian **матриц**, two English equivalents are possible; first, matrix; and second, die (machinist's). The problem would be to select one and only one of these equivalents by a consideration of whether the text material is concerned with the theory of matrices, or a die for stamping out ski parts. This problem is called the nongrammatical incident-meaning problem.

3. By rearrangement of word order to form the output into more conventional English. Inflected languages are especially free as far as word order is concerned, while English, a very uninflected language, has a very rigid word-order form.

The nongrammatical incident-meaning problem and the word-rearrangement problem are not considered in detail in this paper. The nongrammatical incident-meaning problem is an important and difficult problem on which little has been done. One of the more interesting attacks on this problem has been made by a Cambridge group. The word-order problem turns out to be of lesser importance than the other two problems, as the examples show.

This paper is concerned with improvements in the word-for-word translation of unselected Russian-English material by consideration of the grammatical usage of the word in the sentence, as listed previously in item 1. Actually only the adjectives, substantives, and prepositions of the material are processed. Verbs, gerunds,

---

Paper 57-1062, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted April 15, 1957; made available for printing August 15, 1957.

Robert E. Wall, Jr., and Udo K. Niehaus are with the University of Washington, Seattle, Wash.

participles, conjunctions, adverbs, and exclamations are not processed, but are transferred to the output in the input form. Thus this paper is concerned mainly with the location and analysis of noun phrases, and the insertion of the proper English prepositions to render these noun phrases reasonably grammatical in English. Several ways in which the quality of the translation might be measured are also considered briefly.

The Translation Process

The general steps in the translation process have been discussed in detail in previous papers.[1,2] Briefly the translation process consists of four steps:

1. Coding the source-language printed material into machine code.
2. A dictionary search. After this step a word-for-word translation has been obtained.
3. Logical processing of the material to improve the quality of the output over that of the word-for-word translation.
4. Printing out the translation in the target language.

The print-out routine turns out to be a relatively simple problem, while the coding, or reading of the source-language material, is a very difficult problem. Both problems are being considered very actively by various groups.[3,4] Neither of these two problems is considered here since they are somewhat extraneous to the theme of this paper. This discussion begins by consideration of the dictionary search. The layout of the dictionary is that developed by the Far Eastern Department of the University of Washington under a subcontract from the Air Force through the International Telemeter Corporation. This dictionary was prepared for use with the International Telemeter Corporation high-speed photoscopic memory.[5] A typical entry in the dictionary is as follows:

◇@волосяной % (OF)(TO/FOR)(BY/WITH/AS)HAIR/CAPILLARY##

This dictionary was designed for strict word-for-word translation. In the foregoing entry the symbol combination ◇@ is an order to the machine to start to compare the incoming text word with the dictionary entry which follows. The % symbol commands the machine to stop the comparison, and if complete correlation was established between the ◇@ symbols and the % sign, then the machine is to read out everything in the dictionary which is entered between the % symbol and the ## symbols. The brackets indicate words which may or may not be necessary to the translation. The reader is required to determine for himself whether or not the bracketed words are necessary. The words separated by the / sign indicate a choice. Thus either "hair" or "capillary" may be used; the reader must determine this for himself. Since the "HAIR/CAPILLARY" combination is not bracketed, one or the other of these words must be used. Thus the reader has the choice of interpreting this entry in twelve different ways: of hair, to hair, for hair, . . ., of capillary, to capillary, for capillary, . . .

After the dictionary search is completed, a word-for-word translation exists in the machine and may be printed out, or subjected to logical processing in order to improve the output. As an example of a word-for-word translation consider the Russian sentence:

поводом для лечения новокаином нервной импотенции послужило для нас сообщение нескольких больных получавших внутривеные вливания новокаина в связи с гипертонической болезнью, о том, что после проведенного курса вливаний у них исчезли явления преждевременной эякуляции

The word-for-word translation of this sentence would be:

(by/with/as) reason for (of) treatment(s) (of) (to/for) (by/with/as) novacain (of) (to/for) (by/with/as) nerve/nervous (of) (to/for) impotence(s) served for us information/report (of) several/somewhat (of) sick/patients (who) were receiving intravenous (of) injection(s) (of) novacain in/to/at/on/of/like (of) (to/for) connection/communication /bond3 /coupling83 /liaison97 /brace94 from/(from off)/with (of) (to/for) (by/with/as) hypertensive (by/with/as) illness about/against/with that/volume that after/(another time) (of) completed (of) course (of) injections at/by/with/from them disappeared (of) appearance/phenomenon/symptom(s) (of) (to/for) (by/with/as) premature (of) (to/for) ejaculation

In the foregoing translation, the numbers indicate equivalent words which pertain only to a specific field. For instance, 83 refers to the field of electrical engineering.

Logical Grammatical Processing

The necessity of a choice of English prepositions which must be made by the reader is caused by the grammatical ambiguity of the source-language words when they are considered in isolation. For instance, in the case of the in-isolation translation of the Russian adjective волосяной, which was listed in the foregoing, the English preposition "of" denotes that the word might be genitive, and the "to/for" combination indicates that it might be dative. The "by/with/as" entry shows that in addition this word might be instrumental. It is also true that this particular inflection of the word might be used in the nominative, accusative, or prepositional cases. This suggests a different method of storage of the English equivalents for a particular word in the dictionary. If the grammatical information about the source-language word is stored in the form of a linquistic tag in the individual dictionary entries instead of storing all possible English prepositions which are implied by the particular inflection of the source-language word, more information about the word is available to the machine programs than would be available from the existence of the English prepositions. When these tags are used, the English prepositions would be inserted later in their proper places by the machine.

Storing grammatical tags with the individual entries rather than the actual English prepositions gives two benefits: first, the required storage is considerably reduced; and second, the tags allow certain logical processing to be accomplished which considerably improves the quality of the translation.

Storing a linguistic tag instead of the English prepositions actually saves considerable storage space in the dictionary. A count of a few pages of the translation dictionary which is being prepared by the University of Washington machine translation group shows that an average of about 45 bits are required per entry to store the possible English prepositions which are required for translation on a strict word-for-word basis. If a tag is used it is necessary to consider four form classes (adjectives, prepositions, substantives, and none of these) and six cases which may be either singular or plural. Thus $5 + 2(6) = 17$ bits are required to store all the information which is necessary properly to insert the English prepositions in the translated material. Note that no gender information is included since so far it has been found that gender apparently does not allow any improvement in the translation.

The logical processing which is made possible by the use of tags actually turns out to be much more important than the saving in storage. In the example of a word-for-word translation which was presented previously, one must spend considerable time in contemplation of the material before it begins to make sense. Some improvement in the translation must be made if it is to have any degree of general utility. While the information in these 17 bit tags is far from enough to allow translations of any great literary beauty, it will be shown that these tags make possible logical processing which can reduce the number of redundant words in the translation to about half that of the word-for-word translation. The elimination of these redundant words considerably improves the understandability of the output material.

The logical processing starts with the first word of a sentence, and proceeds by considering the tags, and only the tags, which have been obtained from the dictionary. For instance, if a preposition is encountered, the machine notes that a grouping is starting, since the Russian preposition is always the first word of the grouping which it is in. If the next word is an adjective, the machine compares the tag of the preposition with that of the adjective, eliminating all those grammatical usages which either the preposition or the adjective might have in isolation which are not possible when those two words co-occur. The machine then proceeds, comparing tags in this way until a substantive is encountered, which the machine will interpret as the last word of the grouping. If a word is encountered which has no grammatical usage which is compatible with the words considered in the grouping up to that point, the machine will consider that a grouping has been completed even though a substantive has not been encountered, and a new grouping will be considered to start with the nonconforming word. When the machine inserts a preposition and the only possibility for the translation is the case represented by that preposition (or a group of prepositions when the entire group represents a single case), then the machine will enter the preposition without brackets indicating that there is no longer a choice; the preposition must be used.

Several examples of the improvement in output material which may be expected from this simple processing are now presented. First, a processed translation of the sentence which was presented previously is displayed:

by/with/as reason for treatment by/with/as novocain (of) (to/for) nerve/nervous impotence served for us information/report of several/somewhat sick/patients (who) were receiving intravenous injections of novocain in/at connection/communication/bond3/coupling83/liaison97/brace94 from/(from off)/with hypertensive illness about/against/with that/volume that after/(another time) completed course of injections at/by/with/from them disappeared (of) appearance/phenomenon/symptom(s) (of) (to/for) premature ejaculation.

A translation by a polyglot preserving the Russian word order follows:

as a motive for treatment with novocaine of nervous impotence, served for us a report by several patients who were receiving intravenous injections of novocain in connection with hypertensive illness, about the fact that after the carried out course of injections from them disappeared the phenomena of premature ejaculation.

Note that the polyglot, who was not trained in the medical sciences, left out the most desirable equivalent for the Russian явления, which is "symptoms" in this case. Thus a machine translation will always display all possible equivalents while the polyglot, because of unfamiliarity with the special field of the paper, might leave out some of the most important equivalents.

The following is presented as the second example:

о лечении нервной импотенции новокаином

The word-for-word translation would be:

about/against/with medical treatment (of) (to/for) (by/with/as) nerve/nervous (of) (to/for) impotence(s) (by/with/as) novocain

The translation using the program presented in this paper would be:

about medical treatment (of) (to/for) nerve/nervous impotence by/with/as novocain

The third example is as follows:

стенки сосудов пронизаны крупными (диаметр 6–10μ), в более широких до 15μ, окаймленными, округлыми или многоугольными порами с широким окаймлением и с чечевицеобразным горизонтально вытянутым отверстием, не заходящим за окаймление

The word-for-word translation would be:

(of) wall(s) (of) vessels (who/which) (were perforated)/(were threaded) (by/with/as) large/important, diameter 6–10μ, in/to/at/on/of/like more (of) wide vessels (up) to/before 15 μ, (by/with/as) edged (by/with/as) round(ed) or (by/with/as) polygonal (by/with/as) pores/times/seasons from/with/for (by/with/as) wide/extensive (by/with/as) bordering/flange/burr and/but from/with/for (by/with/as) lens-shaped (is) horizontal(ly) (by/with/as) extracted/elongated/exhausted (by/with/as) opening/mesh not (who/which) set(s) (ting) behind/for/after/in/at bordering/flange/burr

The translation using the program presented in this paper would be:

(of) wall(s) of vessels (who/which) (were perforated)/(were threaded) by/with/as large/important, diameter 6–10μ, in/at/on more wide vessels (up) to/before 15μ, by/with/as edged, round(ed) or by/with/as polygonal pores/times/seasons with wide/extension bordering/flange/burr and/but with lens-shaped horizontally extracted/elongated/exhausted opening/mesh not (who/which) set(s) (ting) behind/for bordering/flange/burr

A translation by a polyglot preserving the Russian word order would be:

The walls of the vessels are permeated with large (diameter 6–10μ), in the more wide vessels up to 15μ, bordered, rounded, oval or polygonal pores with a wide border and with a lens-shaped horizontally extended opening not going beyond the border.

One refinement made in the foregoing processed examples which was not mentioned previously is to take advantage of the fact that Russian prepositions which govern more than one case may have the case for a particular application determined accurately by consideration of the case of the adjectives or substantive which the preposition governs. For instance, in the case of the Russian "в," which has in isolation the English equivalents "in/to/at/on/of/like," it is reduced to "in/at/on" in the translation of the third example, since the context of "в" shows that only the prepositional case is possible for this application, while "в" in isolation has the possibilities of either the accusative or prepositional cases.

It should be noted that in some cases the remaining grammatical ambiguities are difficult to resolve even on a semantic basis. For instance, in the second example the only unsolved grammatical problem is that of the combination (of)(to/for). Note that both sentences

about the medical treatment *of* nervous impotence with novocain

and

about the medical treatment *for* nervous impotence with novocain

are grammatical in English. This suggests that just "of" may be used when the word has the possibilities either genitive or dative and no other. Not enough sentences have been checked yet to determine whether this may be done. So far, no examples have been found where it may not.

Flow charts written for the IBM (International Business Machines) *650* computer to perform the indicated operations are shown in Figs. 1 and 2. An example of the input and output of a sentence processed by the computer is shown in Fig. 3. The difference in code is caused by the limitation of the IBM computer. The sequence of events in performing translation on the *650* is as follows:

1. For each source-language word stored in the dictionary at least two IBM cards were prepared: one or more with the English equivalents of the word, and one with the word tag.
2. In translating a particular Russian sen-

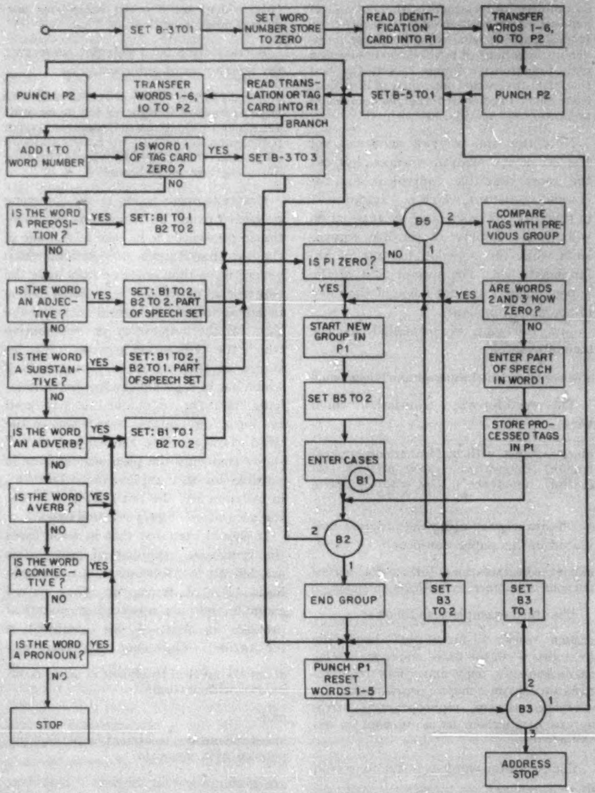

SET B-3 TO 1
SET WORD NUMBER STORE TO ZERO
READ IDENTIFICATION CARD INTO R1
TRANSFER WORDS 1-6, 10 TO P2
PUNCH P2
TRANSFER WORDS 1-6, 10 TO P2
READ TRANSLATION OR TAG CARD INTO R1
BRANCH
SET B-5 TO 1
PUNCH P2
ADD 1 TO WORD NUMBER
IS WORD 1 OF TAG CARD ZERO?
YES
NO
SET B-3 TO 3
IS THE WORD A PREPOSITION?
YES
SET: B1 TO 1 B2 TO 2
NO
B5
2
1
COMPARE TAGS WITH PREVIOUS GROUP
IS P1 ZERO?
NO
YES
IS THE WORD AN ADJECTIVE?
YES
SET: B1 TO 2, B2 TO 2. PART OF SPEECH SET
NO
ARE WORDS 2 AND 3 NOW ZERO?
YES
NO
START NEW GROUP IN P1
IS THE WORD A SUBSTANTIVE?
YES
SET: B1 TO 2, B2 TO 1. PART OF SPEECH SET
NO
ENTER PART OF SPEECH IN WORD 1
SET B5 TO 2
IS THE WORD AN ADVERB?
YES
SET: B1 TO 1 B2 TO 2
NO
STORE PROCESSED TAGS IN P1
ENTER CASES
B1
IS THE WORD A VERB?
YES
NO
2
B2
1
IS THE WORD A CONNECTIVE?
YES
NO
END GROUP
SET B3 TO 2
SET B3 TO 1
IS THE WORD A PRONOUN?
YES
NO
PUNCH P1 RESET WORDS 1-5
2
B3
1
3
STOP
ADDRESS STOP

Fig. 1. Flow chart used to perform indicated operation by an IBM 650 computer

tence, first the cards were pulled for the individual words from the file. The tag cards were run through the computer in text order with the deck for the first round of processing, which is the routine described in this paper.

3. The cards punched out in step 2 were introduced into the computer with a deck which caused the translated sentence to be punched out. The translation was refined from word-for-word by the processing indicated in step 2. The translation includes only those proper English prepositions required by the results of step 2.

4. The cards obtained from step 3 were then tabulated on the accounting machine. The printed-out material was the translation.

Quality of the Translation

The second translation in the foregoing examples is undoubtedly better than the first in each case, but it is difficult to evaluate just how much better. One criterion for judging the improvement would be to count the number of superfluous English words that are present in one translation but not in the other. In answer to this question it may be noted that in the first example the 115 words of the first translation are reduced to 75 words in the second translation; the 21 words of the translation of the second example are reduced to 13 words in the second translation; and the 98 words in the first translation of the third example are reduced to 64 words in the second. In addition, three problems of number are solved in the second translation of the first example.

Another criterion might be to calculate how many sequences may be formed from the words, considering the possible grammatical usages of each word. Thus in the first example, the first word (in Russian) can only be instrumental singular, the second can only be genitive singular or plural, the third word has three possibilities, the fourth has one, and the fifth has four. Thus these five words have: (1)(2)(3)(1)(4) = 24 possible combinations. If this calculation is made for the entire first example, it is found that without any processing there are something like 6.8×10^9 combinations, while with the processing described in this memo, the total number of combinations is reduced to 48!

The two criteria for determining the quality of translation which are presented in the foregoing are attempts to solve the very important question, "How may the quality of a translation be measured?" One way, which is suggested by the first criterion, would be to define a per-cent translation. This might be done by noting the number of problems to be solved in the word-for-word translation of a particular sentence, then noting the number of these problems which are solved by a particular translation. The per-cent translation would then be:

$$T\% = \frac{\text{number of translation problems solved}}{\text{number of translation problems to solve}} \quad (100)$$

In the second example the per-cent translation of the second translation on a grammatical basis would then be:

$$T\% = \frac{9}{11}\,(100) = 82\% \text{ grammatical translation}$$

The 11 problems to be solved are the following:

1. (of)
2. (to/for)
3. (by/with/as)
4. neither (1), (2), nor (3)
5. (of)
6. (to/for)
7. neither (5), nor (6)
8. (s)
9. not (8)
10. (by/with/as)
11. not (10)

The second translation has only one ambiguity, (of) or (to/for), so nine problems have been solved. The second translation would be a 0% nongrammatical translation since none of the nongrammatical problems have been solved.

A similar calculation would give the second translation of the first example a rating of 86% translation, and the third example a rating of 71% translation.

Another measure is suggested by the second criterion. Suppose that a translation ambiguity is defined by the following equation

T_a = log (number of possible sequences that can be formed considering all grammatical possibilities of the words)

This definition of T_a gives the following results:

Example	First Translation	Second Translation
1	$\log(6.8 \times 10^9) = 9.83$	$\log(48) = 1.68$
2	$\log(20) = 1.3$	$\log(2) = 0.3$
3	$\log(49{,}000) = 4.69$	$\log(12) = 1.08$

Still another measure of translation quality would be to use the translation of a sentence by a skilled polyglot as a base, and define a per-cent translation redundancy by the following equation

$$T_r = \frac{(\text{words in translation}) - (\text{words in polyglot translation})}{(\text{words in polyglot translation})} (100)$$

This equation gives the following results for the three examples:

Example	First Translation, Per Cent	Second Translation, Per Cent
1	168	74.5
2	163	62.5
3	180	83

In counting the number of words in the polyglot translation the articles were ignored.

There are many variations that might be made in the definitions for translation quality which were presented, but these criteria do show some interesting results. First of all, the simple grammatical processing which is suggested in this paper solves the bulk of the grammatical problems in a sentence. The remaining problems are essentially nongrammatical in nature. This would be especially true if it turns out to be possible to enter "of" for the "(of)(to/for)" combination as was suggested previously. If this is possible, the second translation of the first example would become a 97% grammatical translation, and the second translation of the second example would become a 100% grammatical translation. The second translation of the third example would remain a 71% translation. This suggests a very interesting application of these measures of the quality of translation. If a routine is included in the machine which will test the translation for quality, then the machine can decide for itself when it has performed enough of a particular type of processing. Thus the first and second examples and would require no further grammatical processing, while the third example is still of poor grammatical quality and should be subjected to further grammatical processing. Either of the criteria defined by $T_{\%}$ or T_a could be built into routines. The criteria defined for T_r could not, however, since the polyglot translation would not be available to the machine for comparison.

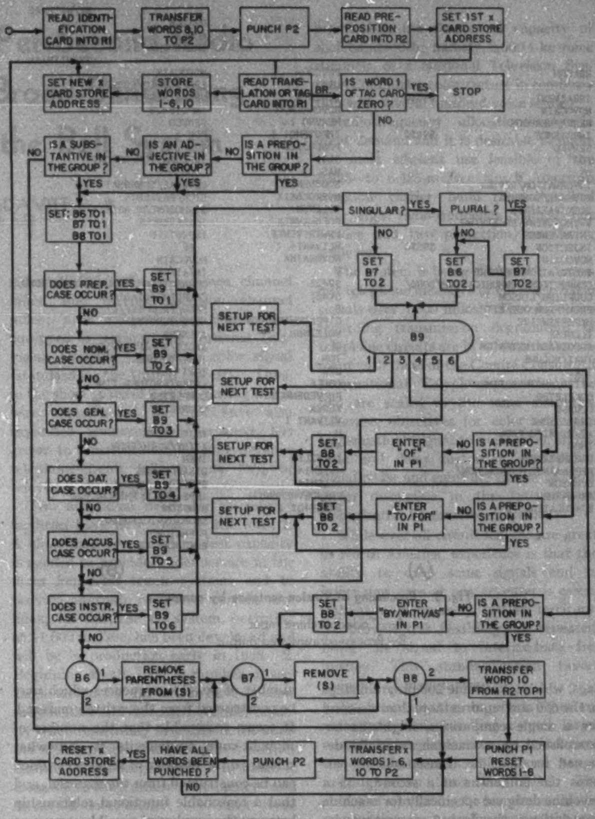

READ IDENTIFICATION CARD INTO R1
TRANSFER WORDS 8,10 TO P2
PUNCH P2
READ PREPOSITION CARD INTO R2
SET 1ST x CARD STORE ADDRESS
SET NEW x CARD STORE ADDRESS
STORE WORDS 1-6, 10
READ TRANSLATION OR TAG CARD INTO R1
BR
IS WORD 1 OF TAG CARD ZERO ?
YES
STOP
NO
IS A SUBSTANTIVE IN THE GROUP ?
IS AN ADJECTIVE IN THE GROUP ?
IS A PREPOSITION IN THE GROUP ?
YES
SET: B6 TO 1 B7 TO 1 B8 TO 1
SINGULAR ?
PLURAL ?
SET B7 TO 2
SET B6 TO 2
SET B7 TO 2
DOES PREP. CASE OCCUR ?
SET B9 TO 1
SETUP FOR NEXT TEST
B9
1 2 3 4 5 6
DOES NOM. CASE OCCUR ?
SET B9 TO 2
DOES GEN. CASE OCCUR ?
SET B9 TO 3
SET B8 TO 2
ENTER "OF" IN P1
IS A PREPOSITION IN THE GROUP?
DOES DAT. CASE OCCUR ?
SET B9 TO 4
ENTER "TO/FOR" IN P1
SET B8 TO 2
DOES ACCUS. CASE OCCUR ?
SET B9 TO 5
DOES INSTR. CASE OCCUR ?
SET B9 TO 6
ENTER "BY/WITH/AS" IN P1
SET B8 TO 2
B6
REMOVE PARENTHESES FROM (S)
B7
REMOVE (S)
B8
TRANSFER WORD 10 FROM R2 TO P1
PUNCH P1 RESET WORDS 1-6
TRANSFER x WORDS 1-6, 10 TO P2
PUNCH P2
HAVE ALL WORDS BEEN PUNCHED ?
RESET x CARD STORE ADDRESS

Fig. 2. Flow chart used to perform indicated operation by an IBM 650 computer

Conclusions

The examples seem to indicate that the bulk of the grammatical incident-meaning problems can be solved by a simple comparison scheme. The processing described in this paper was able to solve about 70% to 80% of the grammatical problems and to simplify tremendously the output. After this simple grammatical processing, the remaining problems are almost entirely nongrammatical in nature. The encouraging thing about this processing routine is that the storage required for the routine and the time required for the computer to run through these routines should be easily acceptable. It is true that it takes from 30 seconds to 1 minute to process a 32- or 40-word Russian sentence on the IBM *650* with this routine, but a computer built especially for machine translation would have some of the operations which take considerable time on the IBM *650* built into logical circuitry in the machine. For instance, the problem of comparing

```
REASON                                     POVODOM
FOR                                        DL4
TREATMENT                      $9S$0       LECENI4
NOVOCAIN                                   NOVOKAINOM
NERVE/NERVOUS                              NERVNO 1
IMPOTENCE                      $9S$0       IMPOTENQI I
SERVED                                     POSLUJILO
FOR                                        DL4
US                                         NAS
INFORMATION/REPORT                         SOO75ENIE
SEVERAL/SOMEWHAT                           NESKOL6KIX
SICK/PATIENTS                              7OL6NYX
$9WHO$OWERE RECEIVING                      POLUCAVWIX
INTRAVENOUS                                VNUTRIVENYE
INJECTION                      $9S$0       VLIVANI4
NOVOCAIN                                   NOVOKAINA
IN/TO/AT/ON/OF/LIKE                        V
CONNECTION/COMMUNICATION/BOND/             SV4ZI
COUPLING/LIASON                            SV4ZI
FROM/FROM OFF/WITH                         S
HYPERTENSIVE                               GIPERTONICESKO1
ILLNESS                                    VOLEZN6H
ABOUT/AGAINST/WITH                         O
THAT/VOLUME                                TOM
THAT                                       CTO
AFTER/ANOTHER TIME                         LOSLE
COMPLETED                                  PROVEDENNOGO
COURSE                                     KURSA
INJECTIONS                                 VLIVANI I
AT/BY/WITH/FROM                            U
THEM                                       NIX
DISAPPEARED                                ISCEZLI
APPEARANCE/PHENOMENON/                     4VLENI4
SYMPTOM                                    4VLENI4
PREMATURE                                  PREJDEVREMENNO1
EJACULATION                                34KLY4QI I
```

(A)

```
                         BY/WITH/AS
REASON
FOR
TREATMENT
                         BY/WITH/AS
NOVOCAIN
$90F$0$0TO/FOR$0
NERVE/NERVOUS
IMPOTENCE
SERVED
FOR
US
INFORMATION/REPORT
  OF
SEVERAL/SOMEWHAT
SICK/PATIENTS
$9WHO$OWERE RECEIVING
INTRAVENOUS
INJECTION                         S
  OF
NOVOCAIN
IN/AT/ON
CONNECTION/COMMUNICATION/BOND/
COUPLING/LIASISON
FROM/FROM OFF/WITH
HYPERTENSIVE
ILLNESS
ABCUT/AGAINST/WITH
THAT/VOLUME
THAT
AFTER/ANOTHER TIME
COMPLETED
COURSE
  OF
INJECTIONS
AT/BY/WITH/FROM
THEM
DISAPPEARED
$90F$0
APPEARANCE/PHENOMENON/
SYMPTON
$90F$0$9TO/FOR$0
PREMATURE
EJACULATION
```

(B)

Fig. 3. Processing of Russian sentence by computer

A—Coded machine input
B—Processed translated output

tags, which takes some 200 program steps in the *650* and requires more than a second for a single comparison, might be accomplished on a machine especially designed for machine translations in perhaps ten millionths of a second. In a machine designed specifically for machine translations, the logical processing described in this paper should take a negligible amount of time compared with the time required for the dictionary search.

The measures of translation quality which were proposed may suggest some ideas whereby one can judge just how good a translation is, and more important, so the machine can measure how good a translation is. The third of these measures would be the least ambiguous to calculate, but would be useless for a routine for the machine. The logarithm criterion T_n might be the more realistic of the three. The amount of mental gymnastics that the brain must go through in coding written material into the brain "thought code" very likely is some function of the number of possible sequences which may be constructed from the written material. It seems reasonable that the number of thought-code equivalents is somewhat similar to the number of sequences which can be constructed from the material, and that a reasonable functional relationship between the number of possible sequences and a person's ability to understand would be a logarithmic function. Thus in the first example, the word-for-word translation probably would be about four times as hard to understand as the second translation, and in the third example the first translation would be about five times as hard to understand as the second translation. This suggests some tests with people to determine just what this functional dependence really is. There would be little sense in spending much time or money in improving the second translation of the second example. The expression is readily understandable in its present form.

Finally, it is interesting to make a comparison between the translation problem and Shannon's chess problem. Shannon's effort is mainly concerned with an attempt to get a computer to play chess—not by storing programs which are triggered by the positions of the pieces on the board alone, but by getting the computer to "learn" to play the game, i.e., the computer will play a better game after it has played 50 games than it played on the first game. Fundamental to any such attempt are some criteria whereby the machine will be able to determine when it has made a good move, or series of moves, and when the moves were bad. With the chess problem this is extremely difficult. The entire object of the game is, "take the opponent's king." If one fails to do this before the opponent accomplishes this same thing, one has played a "bad" game. But just where did the loser go wrong? Everyone who has played chess is familiar with the beginner who very cleverly contrives to take his opponent's individual pieces, but loses sight of the fact that the object is just to take the opponent's king. Often a player will build up an imposing lead in points, then suddenly will lose the game. The chess-playing problem is then an extremely difficult problem for the computer since the quality of individual moves are so hard to measure. The translation problem is easier than the chess problem in one sense: A measure of translation quality can be obtained on a sentence basis. Thus perhaps on an elementary basis the computer can modify its progress to optimize the output material. At least this can be done for the grammatical processing to determine when the processing has proceeded far enough.

References

1. SOME OF THE ENGINEERING ASPECTS OF THE MACHINE TRANSLATION OF LANGUAGE, Robert E. Wall, Jr. *AIEE Transactions*, vol. 75, pt. I, Nov. 1956, pp. 580–85.

2. ENGINEERING PROGRESS IN MACHINE TRANSLATION, R. E. Wall. *The Trend in Engineering at the University of Washington*, Seattle, Wash., vol. 8, no. 3, July 1956, pp. 9–15.

3. PHOTOELECTRIC READER FEEDS BUSINESS MACHINES, David H. Shepard, Clyde C. Heasly, Jr. *Electronics*, New York, N. Y., vol. 28, no. 5, May 1955, pp. 134–37.

4. AN ELECTRON TUBE FOR HIGH SPEED TELEPRINTING, W. H. Bliss, J. E. Ruedy. *RCA Review*, Princeton, N. J., vol. XVI, no. 1, Mar. 1955, pp. 5–12.

5. A HIGH SPEED PHOTOSCOPIC MEMORY, I. Wieselman. *Paper no. 28*, Association for Computing Machinery, Los Angeles, Calif., Aug. 27–29, 1956. (Available from author, International Telemeter Corporation, Los Angeles, Calif.)

6,000-Megacycle-Per-Second Radio Relay System for Broad-Band Long-Haul Service in the Bell System

M. B. McDAVITT
FELLOW AIEE

DEVELOPMENT of microwave communication facilities for Bell System use was undertaken in 1945, taking advantage of prewar research at Bell Telephone Laboratories and of the impetus in technology achieved during the war years. In 1947, an experimental 4,000-mc/sec (megacycle per second) system of moderate bandwidth (*TD-X*)[1] was established between New York, N. Y., and Boston, Mass., a route distance of approximately 220 miles (354 kilometers). It was used commercially for transmission of television signals and experimentally for telephone transmission.

In September 1950 the first *TD-2* system,[2] manufactured commercially by the Western Electric Company, was placed in operation between New York, N. Y., and Chicago, Ill., and was extended a year later to the Pacific Coast, a distance of more than 3,000 miles. It now provides six broad-band radio channels (five service and one protection) in the 3,700- to 4,200-mc/sec band; it has capacity for 600 message channels spaced 4 kc apart (hereinafter called 4-kc channels) or one television signal on each radio-frequency channel; and is designed for transmission over a maximum length of 4,000 miles (6,400 kilometers). It uses a planar triode,[3] producing 1/2 watt of output power, and a metal delay-lens antenna,[4] achieving a gain of 39 db (decibels), horizontal beam width of 2 degrees between half-power points, at 4,000 mc/sec.

Since 1950 applications of the *TD-2* system have grown at a rapid pace, so that today it spans and crisscrosses the United States and Canada, as indicated in Fig. 1. The extent of its use is illustrated by the following statistics: 25,000 route miles, 143,000 broad-band channel miles, 10,500,000 4-kc message channel miles, and 60,000 television channel miles; most of the television channels are equipped to transmit the National Television Systems Committee color signal (standard for the United States[5]). Thus, in the short space of 7 years, during which wire transmission facilities have also expanded enormously, microwave has grown to constitute 22% of long-distance telephone-circuit mileage and 78% of television-circuit mileage.

Two additional needs for microwave facilities have now come to the fore: 1. A short-haul system of modest capacity is needed to provide feeder service to the main long-haul trunk systems and to provide alternate routes for certain wire lines; for this, the *TJ* system, operating at 11,000 mc/sec, has been developed and will be in production early in 1958. 2. Beginning about 1960, a number of sections served by *TD-2* will have become fully loaded and increased channel capacity will be essential; in addition, the number of microwave routes required to enter certain of the larger cities will make it essential to use a different frequency band to avoid interference. To meet these requirements, the *TH*, which is a long-haul broad-band system operating at 6,000 mc/sec, is in the late stage of development and is the subject of this paper.

Objectives of the TH System

In developing the *TH* system, it is appropriate to capitalize, to the maximum extent feasible, on previous experience and on the availability of new techniques and devices, keeping in mind, of course, practical considerations of reasonable development effort, cost of the equipment, and the time when it must be available for service.

A usable bandwidth of 10 mc/sec on each radio-frequency channel is desirable since there is the possibility that any link will have to transmit a color television picture of theatre-screen size. It is desirable to approach, as closely as practicable, the theoretical capacity of such bandwidth; namely, 2,500 4-kc voice channels, or a National Television Systems Committee color picture in combination with several hundred voice signals.

Radio-frequency allocations are in great demand and it is desirable to make the most efficient use feasible of the 5,925- to 6,425-mc/sec North American common carrier band. As described later, eight radio-frequency channels (six service and two protection) are being achieved.

The system is being designed to transmit telephone messages or color television signals over 4,000 miles. The objectives regarding transmission degradation of telephone circuits are in substantial agreement with those of the Comité Consultatif International Telephonique when the two are scaled to the same distances. Likewise, objectives for color television transmission[6] (signal-to-noise ratio, amplitude and delay distortion, and differential phase and gain) are similar to those under discussion in the International Consultative Committee on Radio.

While over-all circuit lengths are great in North America, experience is that the ability to drop some signals and to inject new ones at any repeater is essential, even in sparsely settled territory. To provide this flexibility, all repeaters bring the signal to intermediate frequency. This standardization favors economy of manufacture.

To maintain the required signal-to-noise ratio while handling a much larger number of telephone channels, it is desired to provide considerably more transmitting power than with *TD-2*. The choice of continuing to use planar triodes or adopting the traveling-wave tube was available and was carefully considered. A traveling-wave-tube amplifier with an output power of 5 watts[7] was adopted, mainly because it has wider inherent bandwidth and hence requires fewer adjustments, it injects less delay distortion, it involves less circuit and physical complexity, and it is believed to provide greater system reliability. While estimated costs at comparable stages of manufacturing experience were somewhat favorable to the traveling-wave amplifier, this was not a controlling consideration. If it should prove desirable in the future to design a new 4,000-mc/sec system with greater bandwidth than *TD-2*, it is the present view that a traveling-wave amplifier would be used.

TH will commonly be used to supplement *TD-2* on existing routes; in addition, it will sometimes be desirable for *TJ* to use the same towers. For this

---

Paper 57-1044, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7-11, 1957. Manuscript submitted August 14, 1957; made available for printing August 15, 1957.

M. B. McDAVITT is with Bell Telephone Laboratories, Inc., Murray Hill, N. J.

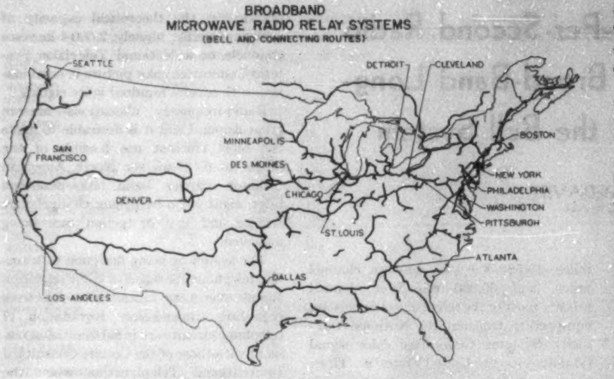

BROADBAND
MICROWAVE RADIO RELAY SYSTEMS
(BELL AND CONNECTING ROUTES)
SEATTLE
SAN FRANCISCO
DENVER
LOS ANGELES
MINNEAPOLIS
DES MOINES
CHICAGO
ST. LOUIS
DALLAS
DETROIT
CLEVELAND
BOSTON
NEW YORK
PHILADELPHIA
WASHINGTON
PITTSBURGH
ATLANTA

Fig. 1. TD-2 microwave relay routes in North America, March 1957

reason, a radio-frequency transmission line and an antenna capable of supporting 4,000, 6,000, and 11,000 mc/sec in any polarization is deemed essential. In anticipation of this, the 2.812-inch (7.13 centimeters) round wave guide and the horn-reflector antenna[8] were put into production in 1955 and have supplanted the rectangular wave guide and delay-lens antenna on all new installations of *TD-2* since then.

Service must be extremely reliable. In addition to telephone and television service, the transmission of information in digital form, such as dialing, teletypewriter signals, and computer data, has assumed a place of major importance in North America. In digital transmission, interruptions of a fraction of a second are likely to cause serious errors. Facilities for fast switching to alternate equipment or channels in case of component failure or fading are provided, as is continuous stand-by power.

A requisite of reliable service is a system that is easy to maintain. Adjustments required by maintenance personnel have been eliminated where possible. Access to the equipment and ease of replacing components have been given detailed attention. Likewise, several new items of test equipment to facilitate maintenance are being developed.

The *TH* system will include a microwave radio order wire and alarm channel. This will avoid the necessity for a paralleling wire line which had proved difficult to provide for *TD-2* in some cases.

TH-System Features

The basic elements of the *TH* system are indicated by the heavily bordered blocks in Fig. 2. Telephone signals will be applied and removed by single sideband amplitude modulation (AM) multiplex equipment of the type used in the *L3* coaxial-cable system.[9] Initially, when used entirely for telephone service, 1,860 4-kc channels will be derived with the frequency allocation shown in Fig. 3. There is a possibility of increasing this loading to 2,220 channels ultimately; if this proves desirable, the fourth master group can be added to the multiplex equipment by conventional means. Fig. 3 also shows the allocation of *TH* frequency space for combined television and telephone transmission; in this case, a gap is left in the telephone band at 7.2 mc/sec to avoid interference from the second harmonic of the strong television color subcarrier signal.

AM-FM CONVERTER

At the last central office preceding the transmitting terminal, newly developed equipment converts the AM signal to a frequency modulated (FM) signal at the intermediate frequency (i-f) of 74.1 mc/sec employed in the radio system. The peak frequency deviation is ± 4 mc/sec. The 10-mc/sec base-band signal produces first-order side bands extending over a 20-mc/sec band, and second-order side bands over an additional 20 mc/sec. The *TH* repeater is gain-equalized over a 28-mc-sec band. While some distortion occurs as a result of not transmitting all of the second-order side bands, the calculated noise from this source in a 4,000-mile system constitutes less than 1% of the total noise power occurring under nonfading conditions.

The elements of the AM-FM transmitting terminal are shown in Fig. 4. The amplified incoming signal applied to the modulating klystron produces an FM signal centered at 6,174.1 mc/sec. This signal and that from the beating klystron oscillating at 6,100 mc/sec, applied to the silicon-crystal converter, produce the 74.1-mc/sec output signal previously mentioned. Both klystrons are of a newly developed type providing unusual linearity and a wide electronic tuning range. High stability of the frequency difference between the two klystrons is maintained by constantly comparing the converter

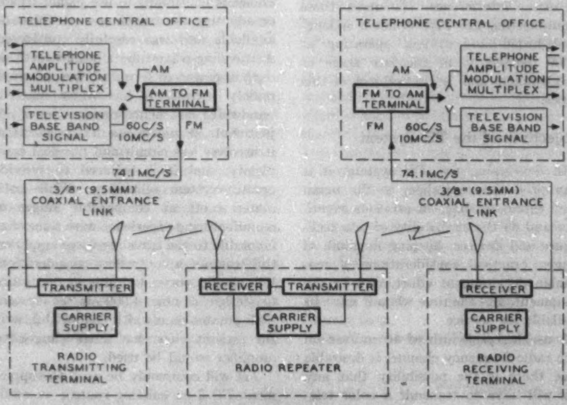

TELEPHONE CENTRAL OFFICE
TELEPHONE AMPLITUDE MODULATION MULTIPLEX
AM
AM TO FM TERMINAL
TELEVISION BASE BAND SIGNAL
60C/S
10MC/S
FM
74.1 MC/S
3/8" (9.5MM) COAXIAL ENTRANCE LINK
TRANSMITTER
CARRIER SUPPLY
RADIO TRANSMITTING TERMINAL
RECEIVER
TRANSMITTER
CARRIER SUPPLY
RADIO REPEATER
TELEPHONE CENTRAL OFFICE
AM
FM TO AM TERMINAL
TELEPHONE AMPLITUDE MODULATION MULTIPLEX
FM
60C/S
10MC/S
TELEVISION BASE BAND SIGNAL
74.1 MC/S
3/8" (9.5MM) COAXIAL ENTRANCE LINK
RECEIVER
CARRIER SUPPLY
RADIO RECEIVING TERMINAL

Fig. 2. TH radio system block diagram

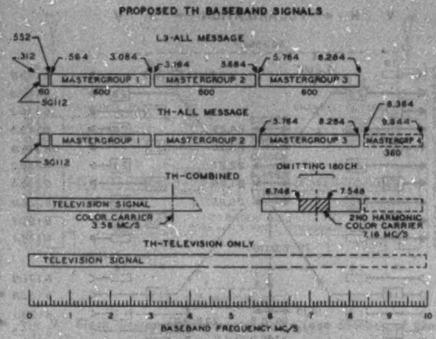

PROPOSED TH BASEBAND SIGNALS
L3-ALL MESSAGE
.552
.312
.564
3.084
3.164
3.684
5.764
8.284
MASTERGROUP 1
MASTERGROUP 2
MASTERGROUP 3
60
600
600
600
SG112
TH-ALL MESSAGE
5.764
8.284
8.364
9.844
MASTERGROUP 1
MASTERGROUP 2
MASTERGROUP 3
MASTERGRP 4
360
SG112
OMITTING 180CH
TH-COMBINED
6.748
7.548
TELEVISION SIGNAL
COLOR CARRIER
3.58 MC/S
2ND HARMONIC
COLOR CARRIER
7.16 MC/S
TH-TELEVISION ONLY
TELEVISION SIGNAL
0
1
2
3
4
5
6
7
8
9
10
BASEBAND FREQUENCY MC/S

Fig. 3. Combination of telephone message channels and television signals for transmission over the TH system

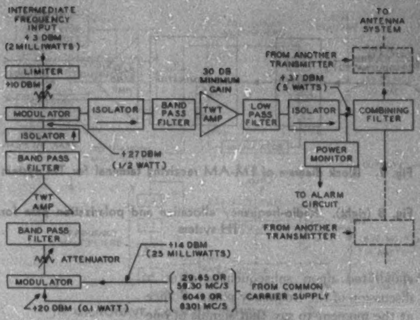

INTERMEDIATE
FREQUENCY
INPUT
+3 DBM
(2 MILLIWATTS)
LIMITER
+10 DBM
MODULATOR
ISOLATOR
BAND PASS FILTER
TWT AMP
BAND PASS FILTER
ATTENUATOR
MODULATOR
+20 DBM (0.1 WATT)
+27 DBM
(1/2 WATT)
ISOLATOR
BAND PASS FILTER
30 DB
MINIMUM
GAIN
TWT AMP
LOW PASS FILTER
ISOLATOR
+37 DBM
(5 WATTS)
COMBINING FILTER
POWER MONITOR
TO ALARM
CIRCUIT
TO
ANTENNA
SYSTEM
FROM ANOTHER
TRANSMITTER
FROM ANOTHER
TRANSMITTER
+14 DBM
(25 MILLIWATTS)
29.65 OR
59.30 MC/S
6049 OR
6301 MC/S
FROM COMMON
CARRIER SUPPLY

Fig. 5. Block diagram of TH radio transmitter

output with a crystal-controlled 74.1-mc/sec oscillator. A voltage proportional to the frequency difference between the converter output and the reference oscillator is derived by means of a limiter-discriminator and synchronous detector. This voltage is applied to the repeller of the beating klystron to adjust its frequency so that the average output frequency of the converter is held close to 74.1 mc/sec.

Entrance Link

The output of the AM-FM terminal is transmitted over 3/8-inch (0.95-centimeter) disk insulated coaxial cable. In most instances, the radio terminal is in the same central-office building as the AM-FM terminal. Occasionally, however, the two are separated at some distance; when this is one mile or less, no repeater is necessary; for greater distances, repeaters similar to, but providing less gain than, the radio i-f amplifier will be used. Equalization is, of course, required in all cases.

Microwave Transmitter

Fig. 5 is a block diagram of the microwave transmitter. The incoming signal first passes through a limiter, which is provided not only at the transmitting terminal, but at every repeater. This is done because it has been found essential to eliminate any amplitude modulation which would be converted to phase modulation by the traveling-wave amplifier and appear as distortion that cannot be compensated by equalization. The amplified and limited signal is fed to the transmitting modulator, which is a gold-bonded germanium diode of new design. While this modulator is capable of giving gain,[10] it is operated in *TH* at a low bias to give uniform impedance over the i-f range, and, in practice, gives neither gain nor loss. It is believed that this is the first application in a system of a diode modulator in which there is no inherent loss.

The microwave carrier signal that is fed to the modulator is discussed later. The output signal from the modulator at transmitting frequency is passed through a microwave ferrite isolator that inserts 0.25-db forward loss, but attenuates signals reflected from subsequent impedance mismatches by 27 db. After passing through a band-pass filter centered at the transmitting frequency, the signal is impressed on the traveling-wave-tube amplifier, about which more is discussed later. In addition to a low-pass filter to suppress harmonics, a second isolator, identical to the one mentioned in the foregoing, is interposed between the amplifier and the antenna transmission line to minimize reflections into the traveling-wave tube from impedance mismatches that cannot be avoided in practice.

The microwave carrier signal, previously referred to, arrives at the transmitter from a common supply, which is

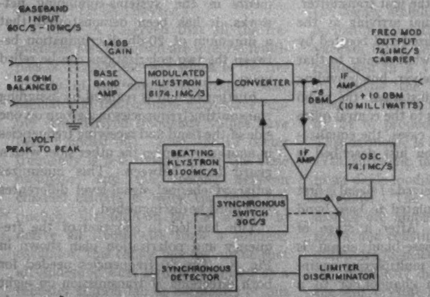

BASEBAND
INPUT
60C/S - 10MC/S
124 OHM
BALANCED
1 VOLT
PEAK TO PEAK
14 DB
GAIN
BASE
BAND
AMP
MODULATED
KLYSTRON
6174.1MC/S
CONVERTER
-8
DBM
IF
AMP
FREQ MOD
OUTPUT
74.1MC/S
CARRIER
+10 DBM
(10 MILLIWATTS)
BEATING
KLYSTRON
6100MC/S
IF
AMP
OSC
74.1MC/S
SYNCHRONOUS
SWITCH
30C/S
SYNCHRONOUS
DETECTOR
LIMITER
DISCRIMINATOR

Fig. 4. Block diagram of AM-FM transmitting terminal for TH system

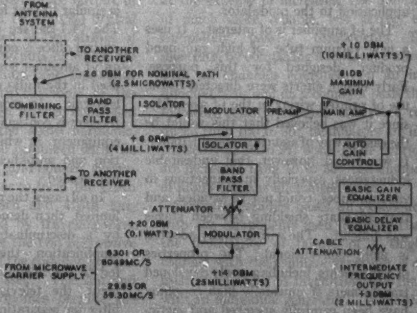

FROM
ANTENNA
SYSTEM
TO ANOTHER
RECEIVER
-26 DBM FOR NOMINAL PATH
(2.5 MICROWATTS)
COMBINING
FILTER
BAND
PASS
FILTER
ISOLATOR
MODULATOR
IF
PREAMP
+6 DBM
(4 MILLIWATTS)
ISOLATOR
TO ANOTHER
RECEIVER
BAND
PASS
FILTER
ATTENUATOR
+20 DBM
(0.1 WATT)
MODULATOR
6301 OR
6049MC/S
FROM MICROWAVE
CARRIER SUPPLY
29.65 OR
59.30MC/S
+14 DBM
(25MILLIWATTS)
+10 DBM
(10MILLIWATTS)
61DB
MAXIMUM
GAIN
IF
MAIN AMP
AUTO
GAIN
CONTROL
BASIC GAIN
EQUALIZER
BASIC DELAY
EQUALIZER
CABLE
ATTENUATION
INTERMEDIATE
FREQUENCY
OUTPUT
+3 DBM
(2 MILLIWATTS)

Fig. 6. Block diagram of TH radio receiver

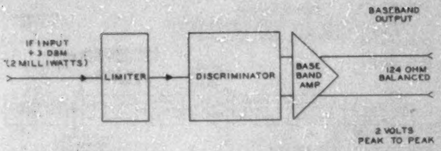
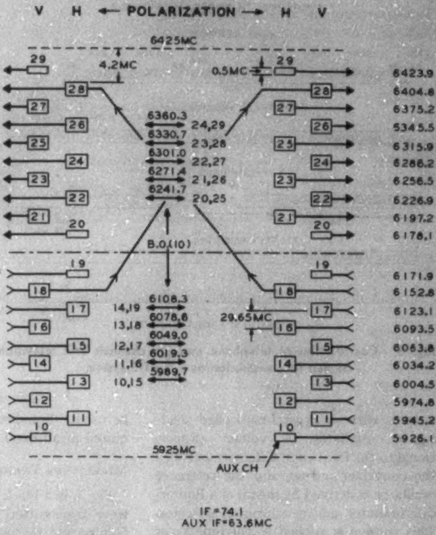


Fig. 7. Block diagram of FM-AM receiving terminal for TH system

Fig. 8 (right). Radio-frequency allocation and polarization plan for TH system



elaborated upon subsequently after a discussion of the frequency plan. Suffice at the moment to say that signals of one of two microwave frequencies and one of two very high frequencies are impressed on a gold-bonded germanium modulator individual to the transmitter to produce the desired frequency. With appropriate side-band filtering, the modulator output is amplified by a traveling-wave tube and thence, through another isolator, is applied to the transmitting modulator.

MICROWAVE RECEIVER

Elements of the *TH* receiver are shown schematically in Fig. 6. After filtering, the incoming signal is passed through an isolator, which absorbs the frequencies, most importantly the image frequency, generated by the receiving modulator that follows. This modulator employs silicon diodes. While slightly inferior to a low-noise traveling-wave tube in noise performance, this arrangement operates the modulator at the lowest possible level and thus minimizes interference due to cross-modulation in the modulator with frequencies from the adjacent channels. The beat-frequency signal is derived from the common microwave carrier supply in the same manner as for the transmitter; here, however, it is not necessary to amplify the beat-frequency signal before application to the modulator.

The i-f amplifier is of interest. It uses a new vacuum tube, of high gain-band product, designed for this purpose. Early design efforts were directed to the use of printed wiring in the i-f assembly, since it was believed that this would be the easiest way of assuring reproducibility. It was found, however, that undesirable inductances, especially in connections to vacuum tubes, could not be avoided, and that adequate shielding between stages was difficult. The wiring finally adopted is more conventional. The properties of components, including newly developed transformers of unusually high coupling, have been carefully specified and meticulous attention has been given to layout. The circuitry is relatively conventional. Thus, a 7-stage amplifier that requires neither factory nor field adjustment has been achieved. Maximum gain is 61 db, with an automatic gain control range of 35 db. The combined passband of the preamplifier and main amplifier, between half-power points, is approximately 50 to 98 mc/sec.

RECEIVING ENTRANCE LINK AND FM-AM TERMINAL

To complete the description of the principal new transmission functional elements, see Fig. 7. The entrance link from the last receiver to the central office is similar to that for the first transmitter. The 74.1-mc/sec signal arriving at the central office is equalized and boosted by an i-f amplifier somewhat similar to that used in the radio receiver, but providing less gain. In most instances, the radio receiver is contiguous to the central-office equipment, in which cases, the equalization and amplification just described is not needed.

In all cases, the received i-f signal is first limited, then demodulated to base band by a discriminator. After a stage of amplification, the base-band signal is fed to the telephone multiplexing equipment, the television channel, or to a filter for separating the telephone signal from the television signal.

RADIO FREQUENCY AND POLARIZATION PLAN

As previously mentioned, in the *TH* system every effort has been made to utilize the available portion of the radio-frequency spectrum with maximum efficiency. It has proved feasible to derive eight broad-band channels and two narrow-band auxiliary channels within the 500-mc/sec allocation.

A major contributing factor is the use of the horn-reflector antenna and round wave-guide transmission line, alluded to previously, which handle vertically and horizontally polarized signals equally well. By the use of polarization-sensitive elements in the systems-combining networks, it has been demonstrated that a minimum of 20-db discrimination between the two polarizations can be relied upon.

Further benefit is realized by assigning transmitting frequencies in a group on one side of the band and receiving frequencies in another group on the other side, with a guard band between. This minimizes adjacent channel signal-level differences that must be distinguished.

This has led to adoption of the frequency and polarization plan shown in Fig. 8. The 20 frequencies needed for both directions of transmission on eight broad-band and two auxiliary channels are tabulated in the right-hand column.

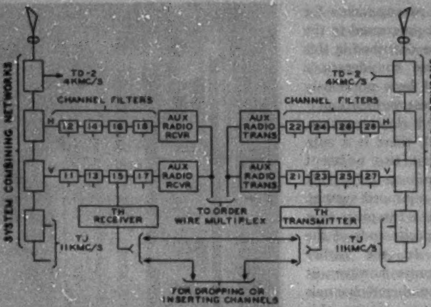

SYSTEM COMBINING NETWORKS
TD-2
4KMC/S
CHANNEL FILTERS
H
12
14
16
18
AUX RADIO RCVR
V
11
13
15
17
AUX RADIO RCVR
TH RECEIVER
TJ
11KMC/S
TO ORDER WIRE MULTIPLEX
AUX RADIO TRANS
22
24
26
28
AUX RADIO TRANS
21
23
25
27
TH TRANSMITTER
FOR DROPPING OR INSERTING CHANNELS

Fig. 9. Block diagram of TH repeater station (one direction of transmission) showing method of combining various receivers and transmitters on common antennas

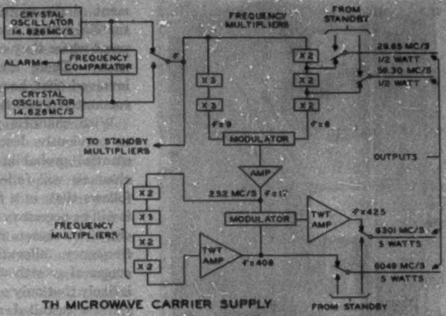

CRYSTAL OSCILLATOR 14.826MC/S
FREQUENCY COMPARATOR
ALARM
CRYSTAL OSCILLATOR 14.826MC/S
f
FREQUENCY MULTIPLIERS
X3
X3
f×9
X2
X2
X2
f×8
MODULATOR
TO STANDBY MULTIPLIERS
FROM STANDBY
29.65 MC/S
1/2 WATT
59.30 MC/S
1/2 WATT
OUTPUTS
AMP
252 MC/S
f×17
FREQUENCY MULTIPLIERS
X2
X3
X2
X2
MODULATOR
TWT AMP
f×425
TWT AMP
f×408
6301 MC/S
5 WATTS
6049 MC/S
5 WATTS
FROM STANDBY
TH MICROWAVE CARRIER SUPPLY

Fig. 10. Block diagram of the common carrier supply for TH system

The channels are given arbitrary designations from 10 to 19 in the lower group, and 20 to 29 in the higher group. Note that the channels are alternately horizontally and vertically polarized. The auxiliary channels are placed at the edges of the two groups since the requirements on their susceptibility to interference are less severe. In Fig. 8, the lower group of frequencies is indicated for receivers, and the upper group for transmitters; it will be apparent that at adjacent repeaters on each side the groups are interchanged.

The center column lists the 10 microwave frequencies generated within the transmitters and receivers, in the manner indicated in the earlier references to Figs. 5 and 6, from the two microwave and two very-high-frequency signals generated in the microwave carrier supply system common to all transmitters and receivers at a station. The upper and lower side bands produced by the i-f modulation of these 10 frequencies constitute the 20 channel frequencies.

(Note that, while the broad-band channel i-f is 74.1 mc/sec, the auxiliary channel i-f is 63.6 mc/sec.)

It may be observed that all the frequencies just discussed are integral multiples of 14.8259 mc/sec, which is the frequency of the closely controlled basic oscillator in the common microwave supply (the derived frequency values have been rounded off). Also, the i-f center frequency is five times that of the basic oscillator, and the spacing between channel centers is twice the basic frequency. In addition to the ability to derive all frequencies from a common oscillator, this choice has the virtue that all undesired beat signals produced at a repeater station fall at harmonics of this oscillator and hence outside the useful frequency band.

It may be seen that, with a 28-mc/sec bandwidth for the main channels and 0.5 mc/sec for the auxiliary channels, the total utilization of the 500-mc/sec allocation is 450 mc/sec. In other words, only 10% of the band is reserved for group separation and for guard bands against channels in the adjoining allocations.

Fig. 9 indicates schematically the arrangement of the systems-combining networks and the channel separation filters for all the receivers and transmitters in a fully equipped repeater station at which *TD-2* and *TJ* systems are also in operation.

Microwave Carrier Supply

Use of a carrier frequency source for the *TH* system common to all transmitters and receivers at one location is a radical departure from the *TD-2* system in which a microwave frequency generator is provided individually in each transmitter and receiver. The basic arrangement of the *TH* carrier supply is shown in Fig. 10. The 14.8259-mc/sec output of the highly stabilized (±0.001%) crystal oscillator is multiplied in two strings of vacuum tubes, the outputs of which are combined in a vacuum-tube modulator to produce the 17th harmonic of the crystal. This is then further multiplied by 24 to produce the 408th harmonic. This harmonic is also combined with the 17th harmonic in a gold-bonded germanium diode modulator to form the 425th harmonic. The 408th and 425th harmonics are amplified by traveling-wave tubes, and constitute the two microwave output frequencies of the common generator, previously discussed. The 2nd and the 4th crystal harmonics provide the very-high-frequency signals, also mentioned earlier.

Because a failure of the common carrier supply would be catastrophic, all elements are provided in duplicate, with appropriate monitoring circuits controlling rapid and reliable automatic switching from regular to stand-by. Coaxial switches are used at the crystal frequency and at its 2nd and 4th harmonics. Newly developed microwave ferrite switches of the Faraday rotation type are used in the microwave arms.

A few words concerning the philosophy of using a common frequency generator are in order. In a repeater station equipped with only one regular and one protection channel, the common supply is slightly more expensive than the individual generators used in *TD-2;* in a fully equipped repeater station, however, the common arrangement is substantially cheaper.

Reliability from the standpoint of component failure is the chief consideration. The common generator achieves a substantial reduction in the number of active components, most prone to failure, at the expense of increasing the number of much more reliable, less active, or completely passive components. To illustrate, by comparing a repeater station with all channels equipped, on an individual basis and on a common basis the following results: vacuum tubes are reduced from 160 to 36; modulators are increased from 16 to 26; wave-guide hybrids are increased from 16 to 30; the number of filters remains unchanged at 48.

Statistical studies based on experience with the components used, indicate that the probability of failure of a single channel with the common generator is at least of an order of magnitude less than with an individual generator. While the probability of failure of all channels individually equipped is almost nonexistent,

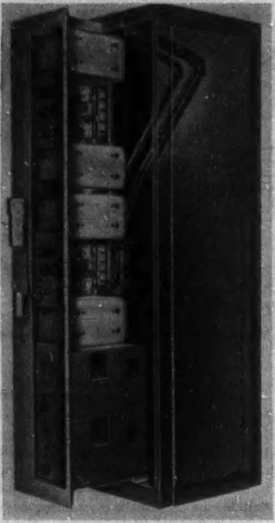


Fig. 11. Bay with two TH radio transmitters and associated high-voltage rectifiers

and that with a common supply in duplicate is finite, the latter is so low as to make the common arrangement highly attractive.

SWITCHING

Automatic switching from a regular to a protection channel is essential to assure service continuity in case of either equipment trouble or fading. The development of the switching arrangements for the *TH* system has not progressed to the point where they can be described in this paper. There are some considerations involved that have not been previously encountered and that deserve mention.

With channels placed close together, as previously described, there is a substantial probability that two adjacent channels will fade simultaneously.[11] It follows that, in a fully equipped system, it will be necessary to have two protection channels, chosen at opposite ends of the frequency allocation. In the earlier stages of growth of a route, however, it is likely that only odd-numbered channels will be installed at first. During this period, one protection channel will serve the purpose.

AUXILIARY RADIO CHANNEL

An auxiliary radio channel to provide voice circuits for maintenance and signaling circuits for controlling switching and for transmission and checking of alarm conditions will be provided in duplicate, as indicated previously. The development of this facility is not sufficiently advanced to permit description at this time.

PHYSICAL ARRANGEMENTS

The physical disposition of the elements of the *TH* system is substantially different from that in *TD-2*. The receiver and transmitter of a channel repeater are located on separate racks. The factory assembly of each receiver and each transmitter includes its channel-separation filter for the appropriate frequency. These are connected in series to the antenna transmission line. The wave-

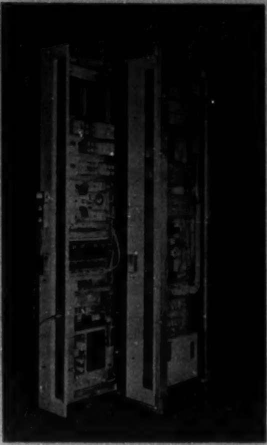


Fig. 12. Common microwave carrier supply bays

guide lines from the microwave carrier supply are simple, since only two frequencies are involved.

There is a preferred, standardized layout for a repeater station that minimizes individual job engineering, simplifies in-

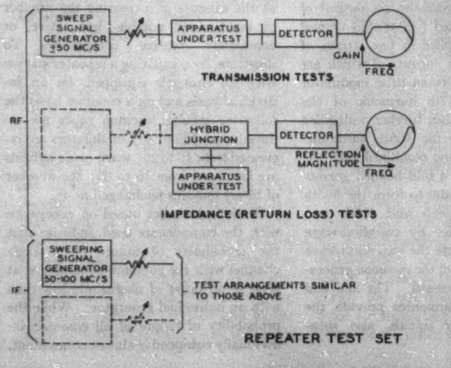

OSCILLOSCOPE
SWEEP SIGNAL GENERATOR ±50 MC/S
APPARATUS UNDER TEST
DETECTOR
GAIN
FREQ
TRANSMISSION TESTS
RF
HYBRID JUNCTION
DETECTOR
REFLECTION MAGNITUDE
FREQ
APPARATUS UNDER TEST
IMPEDANCE (RETURN LOSS) TESTS
SWEEPING SIGNAL GENERATOR 50-100 MC/S
IF
TEST ARRANGEMENTS SIMILAR TO THOSE ABOVE
REPEATER TEST SET

Fig. 13 (left). Block diagram of test equipment for transmission and reflection measurements

Fig. 14 (right). Repeater test console

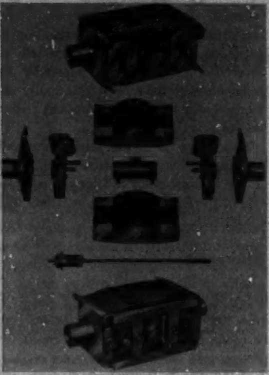


Fig. 15. Assembled and disassembled traveling-wave-tube amplifiers

stallation, and permits orderly predetermined growth of a station that is partially equipped initially. The auxiliary channel bays and the transmitter and receiver bays for the odd-numbered channels (those equipped first) will be in one line; the microwave carrier supply and the transmitter and receiver bays for the even-numbered channels will be in the second line.

A bay 9 feet (2.74 meters) high will accommodate two transmitters or four receivers. By using double-sided racks, the available cubic space can be used to maximum advantage while permitting ready access to all components for maintenance.

Fig. 11 is a rendering of a transmitter bay. (Fig. 11 and subsequent photographs show the latest laboratory models; as a result of experience with these, the appearance of factory-built units will be somewhat different in details.) The double-sided rack is mounted on rollers with the faces of the rack at right angles to the bay line-up. The circuit test points, adjustable elements, fuses, and alarm indicators to which access is needed for routine maintenance are located on the narrow dimension of the rack facing the aisle. A portable plug-in meter avoids the need for several fixed meters. Thus, for day-to-day operation, the rack need not be moved from its normal position.

When internal access is required, the rack is rolled into the aisle space, as shown in Fig. 11. To permit this action, power and miscellaneous low-frequency connections are made to the rack by flexible cables, microwave connections are by specially designed flexible rectangular wave guides, and air for cooling the collectors of the traveling-wave tubes is fed through flexible hose. Thus, ready access is available to both sides of the rack. In addition, rows of cabinets can be mounted against a wall or back to back, thus using the available floor space to the best advantage.

The base of the frame is a casting which supports the superstructure and which also provides a plenum chamber through which the cooling air is fed from a pair of blowers, either of which can supply the entire air requirements.

All *TH* stations will be temperature-controlled to provide the stability of operation needed to meet the exacting system requirements.

Fig. 12 shows the common microwave carrier supply. Three racks are needed: one to house the relatively low frequency gear, shown in the extended position; and the others for the microwave portion. One of the latter is shown in the normal position, but with the bay end-cover removed.

Maintenance Test Equipment

The item of newly developed testing equipment of greatest interest is a portable unit to be provided at every terminal and repeater station for making transmission and impedance measurements at both radio and intermediate frequencies. The principal elements are shown schematically in Fig. 13, and in physical form (with front cover removed) in Fig. 14.

Power Supply

For the supply of reliable direct current, it is not practical to provide storage batteries of the high voltages required for the traveling-wave tubes. All d-c requirements are derived from solid-state rectifiers. Rectifiers are individual to each transmitter and each receiver; hence, the failure of any rectifier will not affect more than one broad-band channel. Likewise, each independent half of the common microwave carrier supply is fed from separate rectifiers.

Alternators, normally driven by a-c motors, also have d-c driving motors powered by a 130-volt battery provided for this purpose only. Thus, a momentary interruption to service is avoided in case the commercial power fails. In the event of an extended power failure, a diesel-engine alternator is started automatically.

New Components

New components used in the *TH* system have been mentioned. A few of the more interesting ones are now discussed.

Traveling-Wave Amplifier

While traveling-wave amplifiers have been used in other microwave systems, the objectives for the one used in *TH* are perhaps more exacting. The tube (Western Electric no. *444A*), displayed in Fig. 15, is designed to cover the band 5,925 to 6,425 mc/sec, to give a minimum gain of 30 db and a power output of 5 watts. The noise figure is 28 db. The anode operates at 2,600 volts, and the helix at 2,400 volts.

Power, cost, and space considerations dictate the use of a permanent magnet, the structure of which is also shown in Fig. 15. The longitudinal focussing field

Fig. 16 (left). Microwave ferrite isolator

Fig. 17 (right). Faraday rotator microwave switch

is a minimum of 600 oersteds, with a maximum allowable net radial field of 2 oersteds. The last limitation places stringent requirements against external magnetic interference. These are achieved with a shield $9\frac{3}{4}$ by 10 by 17 inches (24.8 by 25.4 by 43.2 centimeters), made of cast, low-carbon steel 3/16 inch (4.75 millimeters) thick. The field at the cathode is set at about 15 oersteds to optimize noise performance.

As indicated earlier, a stream of air is passed over the collector to prevent the temperature from rising above 150 degrees centigrade.

Dimensional tolerances are such that tubes can be interchanged in the amplifier with only minor trimming of the magnetic field to optimize focus. The longitudinal positions of the input and output wave guides, and the location of the short-circuiting plungers in these guides are adjustable to provide optimum impedance match for each tube.

MICROWAVE FERRITE ISOLATOR

The isolator has been used extensively in the *TH* design, as indicated in the foregoing system description. A field-displacement type with a single, coated ferrite strip and associated permanent magnet, as shown in Fig. 16, is used throughout. Its insertion loss in the forward direction is 0.25 db or less, and, in the reverse direction, is at least 27 db over the *TH* frequency range.

MICROWAVE FERRITE SWITCH

The switch used to select the regular or spare microwave source in the carrier supply consists of a pair of Faraday rotation devices with electromagnets, illustrated in Fig. 17. They are poled so that, with both magnets de-energized, one transmits toward the common point and the other does not; with both magnets energized, the transmission is interchanged. Thus, "fail-safe" operation is attained. Energy is coupled between the rectangular and cylindrical sections through irises at each end of the latter.

The insertion loss is 0.5 db maximum when transmitting, and 30 db minimum in the nontransmitting condition. The required magnetic field is 35 to 75 oersteds, and the switching time in either direction is 1 millisecond or less.

Note from Fig. 17 that two manually operated shutters, each of which can present a short circuit in the nontransmitting arm, are provided; these are to facilitate removal of either of the switch rotators.

MICROWAVE FERRITE VARIOLOSSER

A third type of interesting new microwave ferrite device is the variolosser (device that inserts controllable variable loss) used in the radio-frequency sweeping signal generator of the repeater test set shown in Fig. 13. The source of power is a klystron oscillator whose frequency is varied ± 50 mc/sec at a 31-cycle/second rate. The power output versus frequency varies rather sharply, whereas a constant power output versus frequency is desired for the sweeper.

The variolosser is a Faraday rotation device whose insertion loss is made to vary in accordance with the current in the associated electromagnet coil. A small portion of the microwave output is rectified to control a d-c amplifier which, in turn, supplies the magnet coil. Thus, there is derived a negative-feedback microwave circuit whose output level is flat to ±0.1 db over ±50 mc/sec and to ±0.05 db over ±25 mc/sec.

Summary

The *TH* system in its entirety is a rather complex aggregation of subsystems. Because of space and time limitations, the foregoing discussion omits many interesting details. These will be covered in future technical papers by the engineers who have been principally responsible for their conception and reduction to practice.

The design is in the final stages and preparation for manufacture is well under way. Present plans call for the completion of the first substantial installation in the latter half of 1959.

References

1. A BROAD BAND MICROWAVE RELAY SYSTEM BETWEEN NEW YORK AND BOSTON, **G. N. Thayer, A. A. Roetken, R. W. Friis, A. L. Durkee.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 37, Feb. 1949, pp. 183–88.

2. THE TD-2 MICROWAVE RADIO RELAY SYSTEM, **A. A. Roetken, K. D. Smith, R. W. Friis.** *Bell System Technical Journal*, New York, N. Y., vol. 30, pt. 2, Oct. 1951, pp. 1041–77.

3. DESIGN FACTORS OF THE BELL TELEPHONE LABORATORIES 1553 TRIODE, **J. A. Morton, R. M. Ryder.** *Ibid.*, vol. 29, Oct. 1950, pp. 496–530.

4. METALLIC DELAY LENSES, **Winston E. Kock.** *Ibid.*, vol. 27, Jan. 1948, pp. 58–82.

5. NTSC COLOR TELEVISION STANDARDS, **D. G. Fink.** *Electronics*, New York, N. Y., vol. 26, no. 12, Dec. 1953, pp. 138–50.

6. INTERCITY B-W AND COLOR TELEVISION TRANSMISSION, **J. M. Barstow.** *Transactions*, Professional Group on Broadcast Transmission Systems, Institute of Radio Engineers, vol. 1, Mar. 1955, pp. 55–64.

7. A MEDIUM POWER TRAVELING-WAVE TUBE FOR 6000-MC RADIO RELAY, **J. P. Laico, H. L. McDowell, C. R. Moster.** *Bell System Technical Journal*, vol. 35, Nov. 1956, pp. 1285–1346. (This paper describes an early developmental version of the tube to be used in TH.)

8. BROAD BAND HORN REFLECTOR ANTENNA, **A. T. Corbin, A. S. May.** *Bell Laboratories Record*, New York, N. Y., vol. 33, Nov. 1955, pp. 401–04.

9. THE L3 COAXIAL SYSTEM—SYSTEM DESIGN, **C. H. Elmendorf, R. D. Ehrbar, R. H. Klie, A. J. Grossman.** *Bell System Technical Journal*, vol. 32, July 1953, pp. 781–832 (see specifically pp. 821–25).

10. TWO-TERMINAL P-N JUNCTION DEVICES FOR FREQUENCY CONVERSION AND COMPUTATION, **Arthur Uhlir, Jr.** *Proceedings*, Institute of Radio Engineers, vol. 44, Sept. 1956, pp. 1183–91.

11. A STATISTICAL STUDY OF FADING OF SUPER-HIGH-FREQUENCY RADIO SIGNALS, **R. L. Kaylor.** *Bell System Technical Journal*, vol. 32, Sept. 1953, pp. 1187–1202.

Kineplex, A Bandwidth-Efficient Binary Transmission System

R. R. MOSIER
NONMEMBER AIEE

R. G. CLABAUGH
ASSOCIATE MEMBER AIEE

SINCE WORLD WAR II, thousands of medium and large-scale scientific and business data-processing computers have been placed in governmental and industrial service throughout the country. With this growth has come the inevitable marriage of computing machinery to wire and radio communications. The need for more reliable and efficient forms of data communications is at the present time greater than ever. Efficient bandwidth utilization and reliability have been major design objectives of the high-capacity Kineplex system. The name *Kineplex* has been adopted because the system employs *kine*matic filters in its detection process and is most frequently employed in multi*plex* configurations. This paper will describe, in brief, the Kineplex system concept. White-noise performance will be compared with an ideal phase-shift system and with actual performance data resulting from tests conducted on a typical narrow-band Frequency Shift (FS) system. Kineplex will be shown, in principle, to make the best possible use of transmission bandwidth.

General System Description

The Kineplex system is a high-capacity phase-shift data-transmission system which will operate in a single telephone voice band of standard quality or equivalent. The system will accept at its input, with suitable conversion, binary data from any source. Specifically, the system is designed to accept either 40 channels of 60–, 75–, or 100–word-per-minute asynchronous telegraph data or, with accessory converting equipment, standard business machine cards at a rate of 200 cards per minute or binary data from high-speed magnetic tapes not to exceed 3,000 bits per second depending on application.

The system has been designed with the technical limitations of the wire and radio media expressly paramount. Radio-frequency (r-f) multipath effects, unequal delays across the passbands of channelizing filters, noisy telephone or radio circuits and bandwidth limitations are factors around which the system was designed. As will be shown, the white-noise performance of the system is superior to standard narrow-band FS by approximately 6 db (decibels). Laboratory tests conducted over both radio and wire circuits have shown the system to be tolerant to most commonly encountered transmission distortion and noise.[1,2]

Fig. 1 is a greatly simplified functional diagram of a 150-bit-per-second, single-tone system. The system represented in the figure might be applicable to the transmission of two 75-bit-per-second channels, or one 150 bit-per-second channel. A synchronizing channel is optionally a part of each system (stable timing generators are otherwise provided). The function of the synchronizing channel is to hold the Kineplex detector in step with the incoming signal pulses. A more detailed block diagram and description of the physical system may be found in reference 1.

Fig. 2 is a 3,000-bit-per-second tone frequency allocation. As shown in the figure, data tones are placed 110 cycles apart from 605 cycles to 2,695 cycles. The synchronizing tone is placed at 2,915 cycles. A bandwidth of approximately 3,400 cycles is required for the transmission of the full 20-tone system; this value includes approximately 500-cycles guard at band extremities.

Modulation, Filtering, and Detection

Phase Modulation

The output of a Kineplex generator is a series of sine-wave pulses. Two subchannels of binary information are encoded into the phase relationship between each pulse and the one preceding. Fig. 3(A) is a representation of this signal.

Filtering

At the Kineplex receiver the transmitted pulse train arrives contaminated by noise and is applied to a pair of time variant or gated kinematic filters.[1] These predetection filters are essentially optimum in the sense that their output signal-to-noise ratios are nearly maximized in the presence of signal and additive white noise.[3] The filters are said to be essentially optimum in that the transfer response of a filter is purposely somewhat broader than the frequency spectrum of the signal. By broadening the response of the filters, tolerance to signal delay distortion has been obtained; such distortion may result from high-frequency multipath and/or the cumulative effects of delay responses in passband devices between transmit and receive terminals; this effect degrades filter output signal-to-noise power ratio by approximately 1.75 db.

The predetection kinematic filters are comprised of Ni-Span C bars to and from which energy is coupled by means of electromagnetic transducers. The bars vibrate in a mode similar to the bars of an xylophone. A typical filter-resonator has a Q in excess of 1,000 and is applied with regeneration such that its time constant is very long with respect to the period during which it is subjected to signal; for practical purposes, the Q of the filter may be considered infinite. The time invariant or ungated transfer response of the filter approximates, therefore, a delta frequency at a signal frequency. In practice, since the filter is gated or keyed such that it is subjected to a signal for a finite time, τ, the response of the filter widens to a $(\sin K\tau)/K\tau$ function.

As described in previous papers,[1,2] two such filters are associated with each Kineplex tone signal. Since the binary information content of a tone is encoded as a phase relationship between each sine-wave pulse and its predecessor, each filter alternately stores as a reference, for one pulse interval, the phase of the signal pulse to which it was last subjected. From the other filter, during this storage interval, energy remaining from a previous pulse is removed after which the filter is subjected to the next signal pulse. Fig. 3(B) describes these filter output characteristics and their relationship to the input signal pulses.

Phase Detection

Phase detection of the two phase-encoded subchannels is accomplished by a pair of product phase detectors, one associated with a subchannel defined as "inphase" and the other with a "quadrature" subchannel. The outputs of the filter pairs are applied directly to these product detectors such that output of the inphase detector is proportional to the scalar product of succeeding pulses; the

---

Paper 57-1042, recommended by the AIEE Communication Theory Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted June 13, 1957; made available for printing August 14, 1957.

R. R. Mosier and R. G. Clabaugh are with the Collins Radio Company, Burbank, Calif.

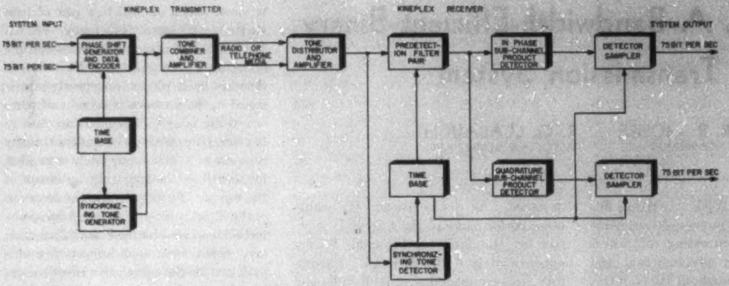

KINEPLEX TRANSMITTER
SYSTEM INPUT
75 BIT PER SEC
75 BIT PER SEC
PHASE SHIFT GENERATOR AND DATA ENCODER
TONE COMBINER AND AMPLIFIER
RADIO OR TELEPHONE MEDIA
TONE DISTRIBUTOR AND AMPLIFIER
TIME BASE
SYNCHRONIZ-ING TONE GENERATOR
KINEPLEX RECEIVER
PREDETECT-ION FILTER PAIR
IN PHASE SUB-CHANNEL PRODUCT DETECTOR
DETECTOR SAMPLER
SYSTEM OUTPUT
75 BIT PER SEC
TIME BASE
QUADRATURE SUB-CHANNEL PRODUCT DETECTOR
DETECTOR SAMPLER
75 BIT PER SEC
SYNCHRONIZ-ING TONE DETECTOR

Fig. 1. Simplified block diagram showing Kineplex concept

quadrature detector output is proportional to the product of succeeding pulses and the sine of the phase difference between succeeding pulses. These product detector transfer characteristics are shown in Fig. 4 relative to the Kineplex generator encoding format.

DETECTOR SAMPLING

The outputs of the product detectors are either positive or negative depending on which of the four possible combinations of binary information (Fig. 4) is encoded in a specific pulse-to-pulse phase relationship. The product detector outputs are sampled at a time, τ, when a filter output signal-to-noise ratio is essentially maximum. Fig. 3(C) shows the sampling interval relative to the filter outputs.

White-Noise Performance

In Fig. 5 are curves of binary error rate versus predetection signal-to-noise ratio for the following systems in the presence of noise: (1) an ideal phase-shift system much like Kineplex but employing a noiseless phase reference, (2) theoretical and actual Kineplex, and (3) measured typical asynchronous narrow-band FS system.[3] All curves are shown on the basis of equivalent signal power-per-channel. Accordingly, an equivalent 150-bit-per-second FS system is assumed. Actual laboratory measurements, obtain from a 75-bit-per-second test on this equipment were, therefore, normalized by the addition of 3 db to all measured values of signal-to-noise power ratio.

IDEAL SYSTEM PERFORMANCE

The ideal phase-shift performance curve, system (1) of the foregoing, of Fig. 5 was computed by a consideration of a system in principle identical to Kineplex but based on phase comparison with a perfect reference; i.e., one not corrupted by noise. The voltage vectors shown in Fig. 6 are included to demonstrate the voltage relationships of an assumed ideal subchannel product detector. Shown in the figure are quadrature and inphase reference vectors, R_q and R_i, the signal, S, and its superposed quadrature and inphase Gaussian noise components, N_{sq} and N_{si}. The output of the quadrature subchannel product detector will be

$$E_o=(R_q)\left(\frac{S}{\sqrt{2}}\right)+(R_q)N_{sq}) \qquad (1)$$

The detector output signal-to-noise power ratio will be

$$\left[\frac{S}{N}\right]_P=\frac{(SR_q)^2}{2N_{sq}^2}=\frac{S^2}{2\sigma^2} \qquad (2)$$

where σ^2 is the mean energy contribution of the quadrature noise distribution and $|S|=|R_q|=$ unity.

The error probability of the ideal phase-shift system is based on the probability of an incorrect binary decision at sampling time at the output of a subchannel product detector. An error will occur at that point when the signal-plus-noise is negative. Fig. 7 represents a Gaussian probability density function distributed about the projection of a constant signal, S, on the perfect reference, R_q. The probability of error for the ideal phase-shift system in the presence of white noise is

$$P_e=\frac{1}{2}\left[1-\mathrm{erf}\left(\frac{S}{\sqrt{2}\sigma}\right)\right] \qquad (3)$$

A plot of this expression is shown in Fig. 5.

KINEPLEX SYSTEM PERFORMANCE

The voltage vectors shown in Fig. 8 demonstrate the noise, signal, and reference voltage relationships for the contaminated reference system, Kineplex. Shown are quadrature and inphase noise terms, N_{rq} and N_{ri}, associated with the transmitted reference; all other terms are as noted in Fig. 5. The output of the assumed ideal scalar product detector of Fig. 8 is given as,

$$E_o=\frac{S}{\sqrt{2}}(R_q+N_{rq}+N_{ri})+R_qN_{sq} \qquad (4)$$

where the product "noise-noise" terms have been neglected. This neglect is valid for error rates below approximately 10%.

The noise amplitudes, N_r, N_{qri}, and N_{sq} are Gaussianly distributed; hence, their

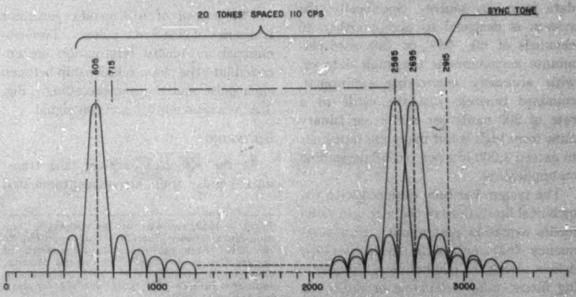


Fig. 2. Tone frequency allocations

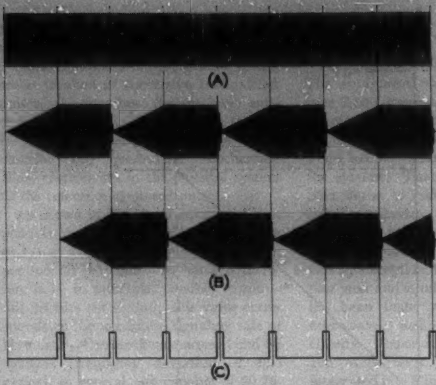


Fig. 3. Filter characteristics and detector sampling

A—Filter input pulse train
B—Amplitude characteristics of a predetection filter pair
C—Detector sampling

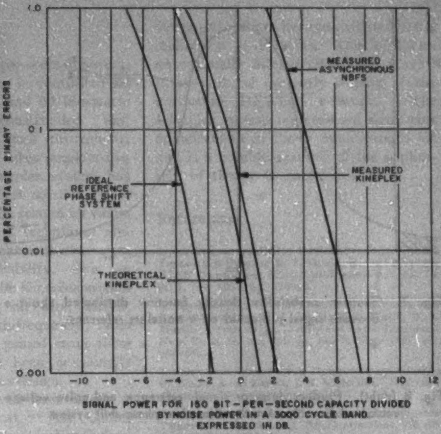


Fig. 5. Comparative white-noise performance

independent energy contributions may be summed. The product detector output signal-to-noise power ratio is, therefore,

$$\left[\frac{S}{N}\right]_p = \frac{S^2R_q^2}{4\sigma^2} = \frac{S^2}{4\sigma^2} \tag{5}$$

where σ^2 is the mean energy contribution of the quadrature and inphase noise distributions and $|S| = |R_q| =$ unity. Consideration of equations 2 and 5 indicate the signal-to-noise power ratio of the perturbed reference case to be approximately 3 db less than that of the ideal system. Intuitively, this is reasoned by noting that with the employment of a noisy reference, essentially 3 db more noise power is added relative to the ideal system. An approximate theoretical error probability expression for the Kineplex phase-shift system is, therefore,

$$P_e = \frac{1}{2}\left[1 - \operatorname{erf}\left(\frac{S}{2\sigma}\right)\right] \tag{6}$$

A plot of equation 6 is shown in Fig. 5. As indicated in the figure, actual measured laboratory performance of Kineplex is, on the average, within 1 db of the

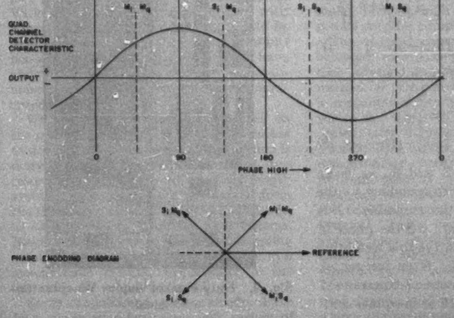

INPHASE CHANNEL DETECTOR CHARACTERISTIC
OUTPUT
+
−
0
90
180
270
0
M_i M_q
S_i M_q
S_i S_q
M_i S_q
QUAD. CHANNEL DETECTOR CHARACTERISTIC
OUTPUT
+
−
0
90
180
270
0
PHASE HIGH
PHASE ENCODING DIAGRAM
S_i M_q
M_i M_q
REFERENCE
S_i S_q
M_i S_q

Fig. 4 (left). Phase relationships in Kineplex data system

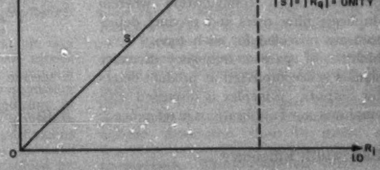

R_q
1.0
N_{sq}
N_{si}
$|S| = |R_q| =$ UNITY
S
0
R_i
1.0

Fig. 6 (above). Product detector signal, reference, and noise voltage vectors relationships for a phase-shift system with a noiseless reference

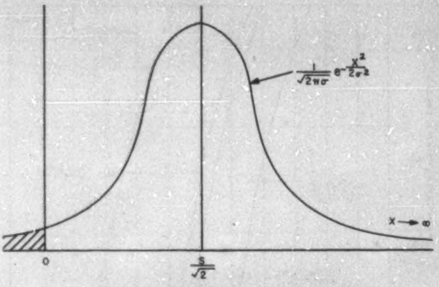


Fig. 7. Gaussian probability density function distributed about a constant signal projected on a noiseless reference

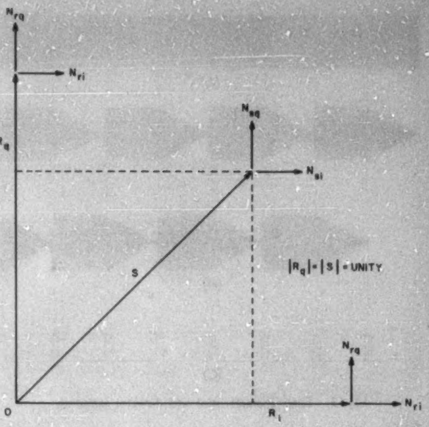


Fig. 8 (right). Product detector signal, reference, and noise voltage vector relationships for the Kineplex phase-shift system

theoretical performance indicated by equation 6. In plotting equations 3 and 6 theoretical performance was degraded approximately 1.75 db in accordance with the provision for tolerance to signal delay distortion.

FS System Performance

The measured white-noise performance of the typical narrow-band FS system was shown in Fig. 5 to be approximately 6 db poorer than Kineplex. It should be pointed out that this value is accurate to the extent of measurements conducted on a limited sampling of typical narrow-band FS systems. The performance difference between these two systems can be attributed, in general, to the optimal nature of Kineplex filtering.

Diversity Performance

For Kineplex r-f applications, typical diversity error rate improvements are obtained by a configuration wherein duplicate filter pairs and product detectors are provided for each tone. Combination of space or frequency diversity signals is accomplished at product-detector outputs. Kineplex is discussed with emphasis on r-f applications in reference 2.

Bandwidth Efficiency

Twenty Kineplex tones are required to transmit the maximum voice band capacity of 3,000 bits per second. As stated, the tones are spaced 110 cycles apart in a bandwidth of approximately 3,400 cycles. In a multitone system, each Kineplex filter pair and associated product detector perform as a correlation detector. The correlation detector associated with a specific tone is subjected to 19 superposed tones, channel noise, and a desired tone.

The Kineplex filters have been discussed in terms of their optimal rejection of noise. It is now intended to establish the orthogonal properties of the tone ensemble and show its integrated products with a detector weighting function to be zero when the arguments of the pulsed tones and a specific weighting function are unequal as described in the following. It will be shown that when these arguments are equal, the integrated product will equal a constant which is proportional to the integration time, τ, and the scalar product of a weighting function and the corresponding pulsed tone of equal argument. The ensemble of 20 pulsed tones may be expressed thusly,

$$S_n(t) = \left\{ S \sum_{n=0}^{19} \cos\left[\left(\frac{2\pi n}{\tau} + \omega_L\right)t + \theta\right]\right\} \{u(t) - u(t-T)\} \quad (7)$$

A specific weighting function is

$$W_i(t) = \left\{ W \cos\left[\frac{2\pi i}{\tau} + \omega_L\right]t\right\} \{u(t) - u(t-\tau)\} \quad (8)$$

where in equations 7 and 8, ω_L is an angular frequency corresponding to the lowest frequency of the ensemble; T is the tone pulse interval; θ is the phase of a tone pulse, $S_i(t)$, relative to a weighting function, $W_i(t)$; and i is an integer, n. The integrated products of equations 7 and 8 may be expressed as in equation 9:

$$\int_0^{\tau} S_n(t) W_i(t) dt = \begin{cases} \frac{SWT}{2} \cos\theta & \text{if } n = i \\ 0 & \text{if } n = i \end{cases} \quad (9)$$

Equation 9 states, in terms of the physical system, that at sampling time, τ, the output of a product-detector will be proportional to the product of the scalar product of successive transmitted pulses and the period over which they are integrated, τ. It further states that at sampling time, no detector output can result from tones which differ from the

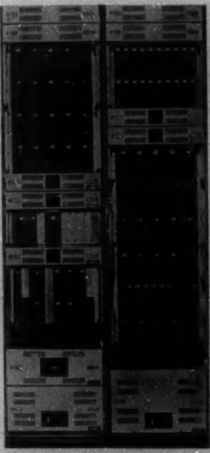


Fig. 9. Forty-channel duplex Kineplex terminal

resonant frequency of a filter-detector by n/τ.

Since equation 9 is proportional to the scalar product of successive transmitted pulses, it describes the inphase subchannel of a tone. If $W_1(t)$ were rotated through 90 degrees, the equation would describe the quadrature subchannel. $W_1(t)$ corresponds in the physical system to the reference or stored pulse.

As shown in the above equations, it is possible to space Kineplex tones as closely as $1/\tau$ cycles with a capacity of $2/\tau$-bits per second per tone. In band B therefore it is, in principle, possible to transmit $2B$ bits per second; this value corresponds to the maximum signaling rate of Nyquist.[4] Conventional narrow-band FS systems (frequency deviation of 42.5 cycles), in general, are capable of 20 75-bit-per-second channels in a single-voice band. Thus Kineplex represents a 2 to 1 improvement in bandwidth efficiency.

Conclusions

Fig. 9 is a 3,000-bit-per-second system. The equipment shown is comprised of over 2,500 transistors and 60 kinematic filters. Modular construction and printed circuitry are used throughout. There are no vacuum tubes employed in the system. The Kineplex system has been shown to make best possible use of frequency spectra. The system by virtue of reliable construction techniques and superior noise performance is characterized by exceptional reliability. Over the past year and a half, the Kineplex system has been under test on wire, carrier, microwave, and high-frequency radio circuits. Experience gained from these and other tests has been continually integrated into Kineplex design.

Kineplex has evolved from nearly a decade of development at the Collins Radio Company. Several earlier Collins-developed systems have contributed to the realization of Kineplex. One of these, an optimally filtered (Predicted Wave[1]) FS system, is currently a part of the northwest DEWline network. The Kineplex system represents a significant advancement in the communications art and is a notable example of the application of theory to useful practice.

References

1. A PREDICTED WAVE-SIGNALLING PHASE-SHIFT TELEGRAPH SYSTEM, E. T. Heald, R. G. Clabaugh. *AIEE Transactions (Electrical Engineering)*, vol. 76, pt. I, July 1957, pp. 316–19.

2. BINARY DATA TRANSMISSION TECHNIQUES FOR LINEAR SYSTEMS, M. L. Doelz, E. T. Heald, D. L. Martin. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 45, no. 5, May 1957, pp. 656–61.

3. OPTIMUM FILTERS FOR DETECTION OF SIGNALS IN NOISE, L. A. Zadeh, J. R. Ragazzini. *Ibid.*, vol. 40, no. 10, Oct. 1952, pp. 1223–31.

4. CERTAIN TOPICS IN TELEGRAPH TRANSMISSION THEORY, H. Nyquist. *AIEE Transactions*, vol. 47, Apr. 1928, pp. 617–44.

◆

Discussion

E. D. Sunde (Bell Telephone Laboratories, Inc., New York, N. Y.): In Kineplex a number of novel and interesting but rather complex techniques have been embodied in the design of a binary pulse system, and an improvement in transmission performance over some of the commercial telegraph systems now in use appears to have been realized. In this paper this is attributed to "the optimal nature of Kineplex filtering." The question raised in this discussion is whether this is in fact the case and whether the performance surpasses that which can be attained with more conventional principles and techniques.

In the paper a comparison is made between Kineplex and a commercial "asynchronous" narrow-band frequency shift system, for equal total signal power and equal total bit-rates in both cases. This comparison indicates about a 6-db advantage in signal-to-noise ratio of Kineplex over the system used for comparison. With a two-fold increase in bit-rate with Kineplex over FS, this advantage would be reduced by 3 db. That is, a 150-bit-per-second Kineplex system would have about a 3-db advantage in signal-to-noise ratio over a 75-bit-per-second asynchronous FS system.

While "asynchronous" is not defined on the paper, it is understood from the context to designate a system in which the timing is derived from start pulses, as in start-stop teleprinters. By contrast in a "synchronous" system the timing would be derived from the signal by electrical "fly-wheel" methods, as used in modern electronic data systems. More timing errors in the pulse sampling would be expected in a start-stop than in a synchronous system, particularly if the system is not properly aligned.

Some unpublished measurements have been made on a frequency-shift system of the same bandwidth (170 cycles) and the same total frequency shift (85 cycles) as the asynchronous system used for comparison in the paper, at 75 bits per second. In the tests the system was operated with synchronous timing, and the error rate was determined as a function of the ratio of signal power to noise power in a 3-kc band. When this ratio is increased by 3 db to allow for a two-fold increase bandwidth and bitrate, the signal-to-noise ratios are smaller by about 1 db than shown in Fig. 5 of the paper for measured Kineplex at 150 bits per second. With a two-fold increase in bit-rate with Kineplex over FS, there would be about a 4-db disadvantage in Kineplex compared to FS. These results are in satisfactory agreement with a theoretical evaluation of the expected performance of synchronous FS compared to Kineplex.

In view of the latter results, I feel that the Kineplex advantage over the asynchronous FS system, cannot be accredited to the "optimal nature of Kineplex filtering" but must be attributed to less than optimum timing techniques in the asynchronous FS system used for comparison.

The channel capacity or total bit rate obtained with Kineplex does not exceed that afforded by the more conventional method of vestigial side-band transmission in conjunction with homodyne detection, in which the signal is applied to a product demodulator together with a synchronous demodulating carrier. Such a synchronous demodulating carrier can be provided by methods used in homodyne detection of vestigial side-band television signals in the television terminals of the *L3* coaxial system.[1] These methods of providing a synchronous demodulating carrier do not entail the inherent 3-db noise penalty incurred by the method used in Kineplex. Moreover, the entire available band can be used directly, and need not be divided into a large number of smaller bands by frequency division, as is necessary with Kineplex if the full potentiality of this method is to be realized. A television channel of 5- to 10-megacycle bandwidth could be used for binary pulse transmission at a bit rate per cycle at least equal to that realized with Kineplex (about 1 bit per cycle) and with some improvement in signal-to-noise ratio. These methods are used in vestigial side-band television channels because they are more economical than simpler and less efficient detection methods that would entail a more expensive transmission medium of either greater bandwidth or shorter repeater spacings. The same methods could be applied to telegraphy or data transmission over voice channels. However, like Kineplex, they require rather complex terminal equipment and in application to narrow-band channels they are not in general as economical as the simpler methods now in common use that are less efficient in their bandwidth utilization. In special situations, such as circuits of great length where extra bandwidth cannot readily be provided and where the signal-to-noise ratio is adequate, such methods of Kineplex might be used to advantage.

In Kineplex two signals are in effect transmitted simultaneously on two carriers of the same frequency but displaced 90 degrees in phase. To derive a carrier for homodyne detection (product demodulation) under this condition is a much more difficult proposition than to derive a demodulating carrier in vestigial side-band transmission. This problem has been overcome in Kineplex by techniques that are quite ingenious though rather complex. However, as regards the net result, in summary my contentions are that:

1. The advantage in signal-to-noise ratio of Kineplex over an "asynchronous" FS system must be attributed to synchronous timing in Kineplex, rather than to the "optimal nature of Kineplex filtering."

2. For equal bit rates, Kineplex would have no advantage in signal-to-noise ratio over an FS system with synchronous timing, and probably a slight disadvantage.

3. A performance both with respect to channel capacity (total bit rate) and signal-to-noise ratio that is equal to or even better than with Kineplex can be realized with more conventional and simpler techniques in commercial use where they are economically advantageous.

REFERENCE

1. TELEVISION TERMINALS, J. W. Rieke, R. S. Graham. *Bell System Technical Journal*, New York, N. Y., July 1953.

R. R. Mosier and R. G. Clabaugh: Although we are generally unable to agree with the contents of Mr. Sunde's discussion, we wish to thank him for his interest in Kineplex.

We are unaltered in the conclusion that (1) our paper correctly establishes the theoretical and measured performance of the Kineplex system, and (2) indicates that which we hold to be a valid, measured comparison with a standard narrow-band asynchronous Frequency Shift system.

We have noted that theoretical and laboratory evidence to support Mr. Sunde's summary paragraphs have not been included in his discussion. Moreover, his first and second summary paragraphs are unclear as to whether he believes there exists an improvement in Kineplex relative to FSK systems.

The discusser refers to an *L3* vestigial side-band television transmission system which was described in part by Rieke and Graham (see reference 1 of the discussion). We believe that neither this reference nor Mr Sunde's discussion establishes the signal-to-noise performance of this system. We are, therefore, without opportunity to compare the system with Kineplex. We feel that a theoretical or laboratory signal-to-noise comparison is necessary to give meaning to a discussion of this kind.

The BIZMAC II Tapefile

J. A. BRUSTMAN
NONMEMBER AIEE

J. AGIN
MEMBER AIEE

THE BIZMAC system was initially designed to meet the requirements of large integrated clerical operations. This was accomplished by analyzing the operations in public utilities, life-insurance companies, government agencies, merchandising and manufacturing firms. As a result of these studies, the BIZMAC I system was designed. Three BIZMAC I systems, with a total number of more than 200 tape stations, are now in service.

To broaden the market, the system has since been critically analyzed for any restrictions on its application to lower volume clerical operation. From this analysis evolved a set of functional enlargements which were incorporated into the BIZMAC II system. These functional enlargements include, among others, the design of an improved tape station with increased pulse packing on magnetic tape and a greater pulse rate. In addition, a tape station is now available for improved manual handling of tape reels.

In the BIZMAC system, the Tapefile is the basic data-storage device. In this capacity, the Tapefile is used for both long-term storage and intermediate working storage.

The BIZMAC magnetic-tape stations and their supporting equipment can be grouped into compact Tapefile assemblies. Such a Tapefile assembly is shown in Fig. 1. Each device is mounted on a pull-out panel. The size of the front panel is 11 inches wide and 30 inches high. The depth of the complete assembly is 42 inches. The basic units of the Tapefile assembly are the tape station, the amplifier control unit, and the power supply. Size and equipment complement are determined by the customer's needs.

Fig. 2 shows the new tape station for improved manual handling. The top section contains the tape-transport mechanism. The lower section contains the read and write amplifiers and the tape control circuits, including the servocontrol unit. Both sections are hinged and may be swung out for service.

Two special controls in this equipment facilitate rapid tape threading and unloading. The cabinet houses two complete stations. The second station is mounted in the back. The size of the cabinet is 3 feet wide, 2 feet deep, and 6 feet high.

Magnetic Recording

The recording on magnetic tape in the BIZMAC system is of the pulsed return-to-zero type, with a 70-kc a-c field for erasing.

The BIZMAC II tape station is designed for recording at a character density of either 125 or 210 characters per linear inch on 5/8-inch-wide magnetic tape. The tape speed for either recording density is 80 inches per second. Thus, the character rates are either 10 kc or 16.7 kc. Approximately 6 million alphanumeric characters can be stored on a standard 2,400-foot tape reel recorded at the higher packing density.

The method of recording BIZMAC characters on magnetic tape is unique. The seven bits of a BIZMAC character are redundantly recorded in 14 channels. Each bit is recorded in two separate and distinct locations on the surface of the magnetic tape. The two recorded dipoles are separated by approximately half of the tape width. Fig. 3 illustrates the pattern which would result from recording a "one" in each of the seven information channels. The dipoles corresponding to any given bit are identified by the information channel number and its primed counterpart. During the reading process, the output from each of the two half bits is added in phase by virtue of internal connections in the magnetic head. Each half-bit contributes 50% of the read signal. The read amplifiers are designed to accept a signal down to 25% of normal output. Therefore, the loss of either half bit in any given information channel does not result in the loss of a character.

The dual-channel recording used in the BIZMAC system permits close packing of tracks without undue crosstalk. As can be seen from the drawing, the crosstalk from adjacent channels is cancelled, with the exception of the two outside channels. Additional shielding between channels 7 and $1'$ is provided to reduce this crosstalk.

Tape-Drive Assembly

Fig. 4 shows the tape-transport mechanism which consists of two magnetic-tape reels, two tape-bin reservoirs, two reel-stripper assemblies, and the head assembly.

A rather unique servo system is used for maintaining a buffer tape storage between the head assembly and the tape reels. This tape buffer consists of a bin-

---

Paper 57-674, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted August 24, 1956; made available for printing May 28, 1957.

J. A. BRUSTMAN and J. AGIN are with the Radio Corporation of America, Camden, N. J.

Fig. 1. Tapefile assembly, pull-out type

type reservoir containing approximately 15 feet of magnetic tape in a balanced condition. A proportional control is derived by sensing the weight of the tape in the bins. A change from normal in the amount of tape resting in the bin displaces the core of a differential transformer, producing an error signal. A rate generator is belt-coupled to the reel motor. Its output, which is proportional to the angular velocity of the tape reel, is connected in series with the output of the differential transformer. The net voltage is applied to the servoamplifier to provide a smooth flow of tape to restore tape balance in the bin.

The tape remains essentially in a tensionless condition throughout the tape path, except at two points near the tape reels. Tension is only introduced to assure smooth winding and unwinding of the tape. During unwinding, tape tension is introduced by a reel-stripper capstan associated with the supply reel. The stripper capstan is coupled to the reel shaft through a belt and an overdriven slip clutch.

When the tape is wound onto the take-up reel, uniform tension is provided by another slip clutch which forces a friction pad against the tape where it passes over a guide post.

Fig. 5 shows the head assembly. The knob located above the tape-guide assembly permits access to the recording area, and at the same time lifts the pressure pads for tape loading. Precise guiding of tape over the magnetic heads is accomplished through the use of two hardened parallel guide plates approximately 6 inches in length. The top of the magnetic heads presents a highly polished metal surface to the magnetic tape. A pressure pad above the read-write head maintains a tape-to-head pressure of approximately 25 grams.

The magnetic heads are fastened to a precisely machined surface and never require realignment. The construction and mounting tolerances of the magnetic heads are such that they permit complete interchangeability of heads between tape stations.

A large number of magnetic heads of similar design installed in BIZMAC I equipment have indicated an operating life expectancy of more than 4,000 hours.

Fig. 5 also shows the symmetrical dual-capstan drive. Two capstans, each driven by a separate 4-pole hysteresis synchronous motor, are constantly rotated in opposite directions. To move the tape in either direction, a pressure roller is pressed against the proper capstan; see Fig. 6. The actuators which accomplish this are constructed along loudspeaker design principles. Carefully designed linkages couple the actuator with the pressure rollers. Two separate actuators are used to drive the tape in either direction. The tape station is capable of responding to 50 start-stop cycles per second.

A third actuator is used as a brake for rapidly stopping the magnetic tape. The brake action is accomplished by clamping the tape between the brake shoe and the highly polished surface of the erase head. The specified start time of the magnetic tape is 5 milliseconds, and the stop time is 3 milliseconds. The actual performance of production tape stations is considerably better.

The amplifier control unit contains all the electronic circuits; see Fig. 7. The seven plug-in units at the left top are the read preamplifiers. Located below are the seven write amplifiers, the erase oscillator, and other control circuitry. At the lower right is the servocontrol chassis, and directly above, the amplifiers for the tape-motion actuators. The control panel at the left is only used for servicing the equipment.

Fig. 8 is a simplified diagram of the read, write, and erase circuitry. Only one of the seven read and write channels is shown. The read signals from the seven dual channels of the magnetic head are transformer-coupled to the inputs of seven preamplifiers. The outputs of these preamplifiers are again transformer-coupled to seven balanced low-impedance lines for transmission to the data-processing machine through the tape trunk. The transmission level is approximately 6 volts peak to peak. Final shaping of the read pulses takes place in the user equipment.

The amplitude of the write pulses received from the data-processing equipment is 10 volts peak to peak. The duration of the write pulses is approximately $7^1/_2$ microseconds. After passing through *AND* gates, the write pulses are clocked out with a 5-microsecond pulse, amplified, and then fed to the

Fig. 2 (left). Manual Tapefile

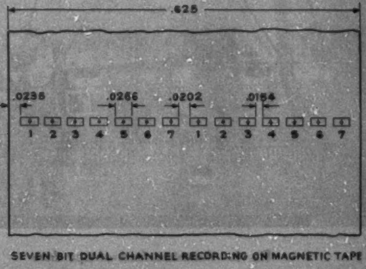


Fig. 3 (right). Recording pattern

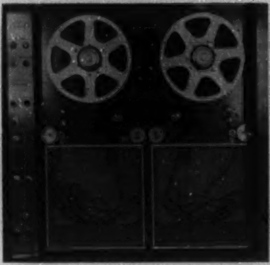


Fig. 4. Complete tape-transport mechanism

magnetic head as current pulses of approximately 60 milliamperes.

The Tapefile trunk contains two sets of information lines. One set of lines transmits the signals from the preamplifier to the user equipment during the read process. During the write process, the same lines provide the return path for the echo signals. The other lines supply the write signals. Telephone-type relays are used for switching. Up to 300 feet of cable can be inserted between the Tapefile and the user equipment.

A simplified block diagram of the tape control circuitry is shown in Fig. 9. Control pulses which cause the tape to move in the forward direction, to stop, or to move in the reverse direction are transmitted over balanced low-impedance lines. These control pulses are stepped up by transformers prior to driving two flip-flops. The two flip-flops provide a memory for the forward-reverse and start-stop conditions necessary for the tape control. The output of the start-stop flip-flop feeds the circuits associated with the drive and brake actuators. After passing through an amplifier, the d-c level from one side of the start-stop flip-flop provides the current for operating the reverse or forward actuators. The output from the other side of the start-stop flip-flop feeds a one-shot delay, which provides a 2-millisecond pulse to the brake drive amplifier.

Fig. 5. Read-write head assembly

The direction of the tape motion is determined by the condition of the forward-reverse flip-flop and its associated relay K. The use of a relay allows time-sharing of the clutch drive circuit and provides a simple and positive interlock against any attempts to drive the tape in two directions at one time. The relay K is of the high-speed mercury-wetted type, which permits changing the tape direction in less than 10 milliseconds.

Data are written onto the tape with respect to two physical markers, namely the beginning-of-tape marker and the end-of-tape warning. A third marker, which is called end-of-tape, is used for internal control in the tape station. The beginning of tape is marked by a window in the magnetic tape which is sensed photo-electrically. This point on the tape is the reference for most Tapefile operations and is the point to which the magnetic tape must be rewound. When the tape is rewound, the beginning-of-tape signal is used to reset the two control flip-flops. This signal is also returned to the user equipment for control purposes. The end-of-tape warning signal is returned to the data-processing equipment as a command to terminate the write-out operation at the next convenient point in the program. The end-of-tape level is used to stop the tape motion.

Fig. 10 shows the power supply, which is housed in a case equivalent in size to four tape-station cabinets of the pull-out type. Regulated voltages are used throughout. Provisions are made to bring up the filament voltage from zero to maximum over a 2-minute period. At the end of this time, nine separate d-c supplies are rippled on, beginning with the most negative voltage. In turning off, the d-c is first rippled off in reverse order, and then the filament voltage is reduced to zero again over a 2-minute period. One unit is capable of supplying power to seven amplifier control units and associated tape stations. Maximum input power to the supply is 7.5 kva. The average line input power is 4 kva.

In most large data-processing systems, there is a need for "work" and "reference" data storage. The work data are often of a temporary nature and must be available for use on short notice. On the other hand, the reference data are usually processed in an orderly sequential manner. As a reference file extends over many tape stations, usually only a small percentage of these is active at any given time. To take advantage of this fact, a variation of the amplifier control unit was designed which can be time-shared by a group of up to five tape stations. The term "multiple

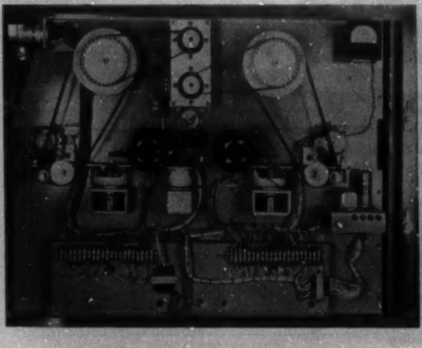


Fig. 6. Actuators, back-of-tape drive

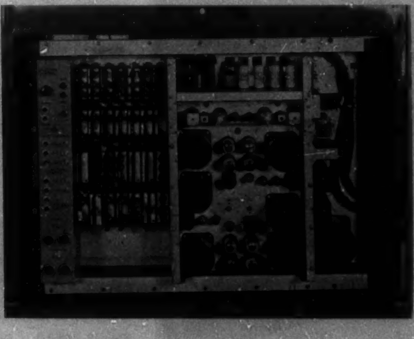


Fig. 7. Amplifier control unit, front view

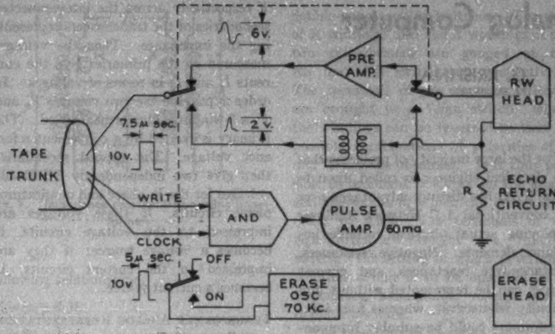

6v.
PRE AMP.
RW HEAD
7.5μ sec.
2 v.
TAPE TRUNK
10v.
ECHO RETURN CIRCUIT
R
WRITE
AND
PULSE AMP.
60ma
CLOCK
5μ sec.
OFF
10v
ON
ERASE OSC 70 Kc.
ERASE HEAD

Fig. 8. Simplified diagram, read, write, and erase signals

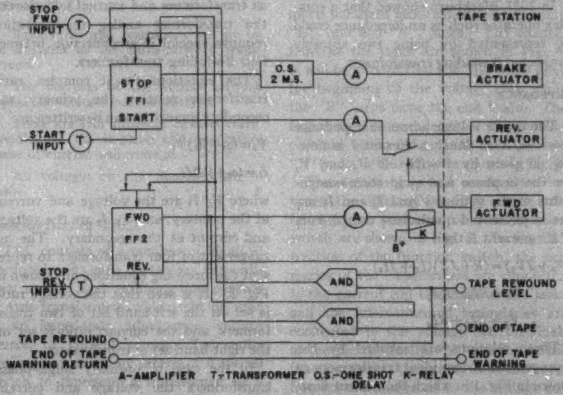

STOP FWD INPUT
T
TAPE STATION
STOP FF1 START
O.S. 2 M.S.
A
BRAKE ACTUATOR
START INPUT
REV. ACTUATOR
FWD FF2 REV.
FWD ACTUATOR
K
B+
STOP REV. INPUT
AND
TAPE REWOUND LEVEL
AND
END OF TAPE
TAPE REWOUND
END OF TAPE WARNING RETURN
END OF TAPE WARNING
A-AMPLIFIER T-TRANSFORMER O.S.-ONE SHOT DELAY K-RELAY

Fig. 9. Simplified diagram, control signals

Fig. 10. Power supply

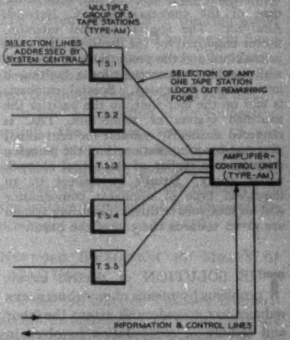

MULTIPLE GROUP OF 5 TAPE STATIONS (TYPE-AM)
SELECTION LINES ADDRESSED BY SYSTEM CENTRAL
T.S.1
SELECTION OF ANY ONE TAPE STATION LOCKS OUT REMAINING FOUR
T.S.2
T.S.3
AMPLIFIER-CONTROL UNIT (TYPE-AM)
T.S.4
T.S.5
INFORMATION & CONTROL LINES

Fig. 11. Block diagram of multiple-group amplifier control unit

group" is applied to such an assembly. The system economies accruing from the use of multiple-tape groups are substantial and permit the retention of reference files on such tape stations. Fig. 11 illustrates the basic block diagram of a multiple group.

A particular tape station within this group of five may be selected by an equipment operator or by a system central operator associated with large BIZMAC installations. When a given tape station is selected, the other stations within the group are locked out. Thus, one trunk to the multiple-type amplifier control unit provides access to up to five tape stations.

Summary

The building-block design of tape stations, together with a manual-type tape station, provide the system engineer with great flexibility in laying out suitable Tapefile assemblies for a variety of equipment complements, floor space, and building requirements. He has at his disposal several types of tape stations—the manual type, which lends itself to rapid changing of magnetic tapes, and the pull-out type, which may either be operated directly with an amplifier control unit or time-shared. Where multiple groups are used, a single power supply may provide power for up to 35 tape stations. The flexibility which has been incorporated into the BIZMAC Tapefiles has proved to be very effective in planning system installations of widely varying requirements. This, in turn, has resulted in significant savings to the customers.

The Transformer Analog Computer

P. VENKATA RAO
ASSOCIATE MEMBER AIEE

G. KRISHNA
NONMEMBER AIEE

Synopsis: The Transformer Analog Computer (TAC) is a very versatile machine capable of solving problems in a-c networks and also linear simultaneous equations with complex coefficients. With the aid of TAC, it is possible to determine the latent roots of matrices and zeros of polynomials, and also the data required to plot the Nyquist diagram of a servosystem.

In this paper, the principle of operation of TAC is explained briefly, and a detailed description of the new double-ended unit developed here is given, along with examples to show how the operation and interconnection of TAC become as simple as in the case of the conventional A-C Network Analyzer. The special parallel interconnector employed for the double-ended units is described and the precautions to be taken while making connections for a multimesh network are indicated. Special emphasis is laid on the fact that, on the basis of the modified system of connection, TAC is rendered eminently suitable for centralized plugging-board operation; and the plugging diagram is a replica of the actual network under study. Details in respect to an improved type of nonlinear compensator and an inexpensive direct measuring system are given towards the end of the paper.

THE SOLUTION of electric circuit problems by means of model networks reduces to a considerable extent the labor and tedium of computation involved in the use of long-hand methods. Among such calculating boards the A-C Network Analyzer has received a high degree of development and it is now widely used for the solution of power system problems. Most of these network analyzers use tapped resistors, inductors, and capacitors which are interconnected to simulate the actual network under study. The A-C Network Analyzer is essentially a steady-state analog analyzer which can be used for studying problems in transients as well by using some special techniques.

TAC is basically a prototype of the Blackburn Network Analyzer.[1,2] Here ideal-voltage transformers are used to represent the resistors, inductors, and capacitors, of the network under study. TAC can be used for the study of A-C networks in the steady state and also for the solution of linear simultaneous equations with constant coefficients, real or complex. The location of the zeros of polynomials and the determination of the latent roots of matrices can also be carried out with the aid of this type of computing equipment.[3]

For the large majority of problems that the electrical engineer is called upon to solve, TAC has distinct advantages over the conventional A-C Network Analyzer employing actual physical variable impedance elements. Negative resistances, resistance-free reactances, and current sources can be represented without any difficulty whatsoever, whereas four ideal transformers could be suitably interconnected to represent symmetrical or asymmetrical mutual coupling effects or phase shifting transformers.

In 1937 Blackburn showed that a complex operator such as an impedance could be represented by using two suitably connected 3-winding transformers.

IMPEDANCE

The vector voltage across an impedance $Z=R+JX$ in which a current I is flowing, is given by $V=ZI$. If V_p and V_q are the in-phase and quadrature components of the voltage V and I_p and I_q are the in-phase and quadrature components of the current I, then

$$(V_p+JV_q)=(R+JX)(I_p+JI_q)$$

$$\therefore V_p=RI_p-XI_q$$

$$V_q=XI_p+RI_q$$

These relations are satisfied by the arrangement of two ideal transformers as shown in Fig. 1 in which the current flowing into the impedance Z is simulated by applying voltages proportional to I_p and I_q, to the primaries of the two transformers. The resistance and reactance of the impedance are the turns-ratio between the secondary and primary of the windings.

ADMITTANCE

The same arrangement of transformers shown in Fig. 1 will represent an admittance $Y=G+JB$ if the current and voltage circuits are interchanged and the tap ratios set as G and B.

SOURCE OF VOLTAGE AND CURRENT

If a voltage V is impressed on an impedance Z, a current I flows through it where $V=ZI$. To simulate this condition in the transformer analog it is necessary to impress voltages V_p and V_q (actually in phase) proportional to the in-phase and quadrature components of V respectively across the interconnected secondaries of the transformers representing the impedance. Then the voltages measured at the primaries give the currents I_p and I_q in terms of voltages. In order to provide the two voltages V_p and V_q a 3-winding transformer is used. The primary is excited by a convenient reference voltage. The tapped secondaries then give two independently adjustable voltages for the in-phase and quadrature-phase circuits. If these voltages are impressed to the voltage circuits, it becomes a voltage source; if they are impressed to the current circuits, it becomes a current source.

TRANSFORMER ANALOG REPRESENTATION OF COUPLING EFFECTS

If the network consists of elements which have coupling between them, such as transformers and mutual impedances, the transformer analog representation requires special interconnection between four 3-winding transformers.

The equations for a complex ratio transformer relating the primary and secondary quantities can be written as:

$$V_2=(a_1+jb_1)V_1$$

$$I_1=(a_2+jb_2)I_2$$

where V_1, I_1 are the voltage and current at the primary, and V_2, I_2 are the voltage and current at the secondary. The arrangement of four transformers to represent the foregoing equations are shown in Fig. 2. It is seen that the voltage ratio is set on the left-hand set of two transformers, and the current ratio is set on the right-hand set of transformers.

In the case of ordinary single-phase transformers the voltage and current ratios are pure numbers, i.e., $a_1=a_2$ and $b_1=b_2=0$. For a phase shifting transformer of voltage ratio 1 to $a+jb$ the current ratio is determined by the fact that power is invariant.

The internal connections between the two sets of transformers to represent mutual impedance between two separate circuits is shown in Fig. 3 giving the

---

Paper 57-1053, recommended by the AIEE Computing Devices Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7-11, 1957. Manuscript submitted March 19, 1957; made available for printing September 24, 1957.

P. VENKATA RAO and G. KRISHNA are with the Indian Institute of Science, Bangalore, India.

The authors wish to thank Prof. G. S. Ghosh for his kind encouragement and constant help during the progress of this work. Thanks are also due to Prof. M. S. Thacker, Director-General, Scientific and Industrial Research, for his interest in this work and for providing adequate funds for the construction of the computer; and to the authorities of the Indian Institute for their permission to publish this paper.

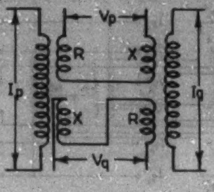


Fig. 1. Impedance representation using transformers

following relationships:

$$V_2=(a_1+jb_1)I_1$$

$$V_1=(a_2+jb_2)I_2$$

The outstanding features of an analyzer employing the transformer analog principle are that:

1. Resistances and reactances are set on the analyzer as the turns-ratio of transformers.
2. The currents of the actual network appear as voltages on the analyzer.
3. There are four different circuits corresponding to the in-phase and quadrature phase of current and voltage.
4. All voltages on the analyzer are in phase.

DESCRIPTION OF TAC

TAC consists of a number of identical universal units each containing two ordinary units. Each ordinary unit consists of two transformers which can be used as an impedance, admittance, voltage source, or current source. Two units are interconnected by means of a special switch to represent a transformer or a mutual impedance.

THE UNIVERSAL UNIT

This composite unit consists of two sections, each having two 3-winding transformers. Each transformer has a primary of 1,000 turns and two tapped secondaries of 999 turns called the *R* and *X* windings. The *R* and *X* windings of each transformer are tapped in a decade arrangement so that turns-ratio from 0.001 to 0.999 are available. These tapped secondaries of the two transformers are brought to two sets of three double-gang tapping switches. The pair of *R* windings and the *X* windings of the two transformers are ganged at the *RG* and *XB* switch sets respectively. The terminals of the tapping switches are brought to a "sign switch" where the windings can be reversed so that all possible combinations of polarities of *R* and *X* may be obtained. It is thus possible to represent $\pm R \pm jX$ on the analyzer, the values of which can be set from 0.001 to 0.999. The terminals of the sign switch are brought to a "function switch," the setting of which determines whether a unit is to be operated as an impedance, admittance, voltage source, or current source.

The terminals of the four circuits of a unit are brought to a plugging board for the purpose of making interconnections between units, by means of two 6-terminal polarized jacks called the beginning "B", and end "E", jacks. The connections at the jacks are shown in Fig. 4. It will be noticed that the "B" jacks have the beginning of the voltage coils while the "E" jacks have the end leads. The beginning and end leads of the current coils are connected to both the "B" and "E" jacks in the sequence shown in Fig. 4.

The details of actual connections between the different windings of the unit for various positions of the function switch are shown in Fig. 5. In the "Z" position of the function switch, the primaries of the transformers are connected to the current coil terminals at the jacks and the interconnected secondaries are connected to the voltage coil terminals at the jacks. In the "Y" position of the function switch, the primaries of the transformers are connected to the voltage coil terminals of the jacks and the secondaries are connected to the current coil terminals. In the "E" and "I" positions of the function switch, the primaries are disconnected from the jack terminals and excitation voltage is applied to the left-hand transformer primary. The *R* and *X* secondary windings of the left-hand transformer are connected to the voltage coil terminals of the jacks in the "E" position and to the current coil terminals in the "I" position.

The terminals of the function switch are brought to a "connection switch." Full details in respect of interconnections between units by the connection switch in the "T" and "M" positions are shown in Fig. 6(A) and 6(B). The two sections of the universal unit are operated as separate complex multiplier units in the "S" position of the connection switch. In the "T" position of the connection switch the two units are connected to act as a transformer unit. In the "T" and "M" positions of the connection switch it is necessary to keep the function switch position of the two units in position "Z". In the "T" position of the connection switch, the voltage terminals of the left-hand unit jacks are connected to the primaries of the left-hand unit, and the voltage terminals of the right-hand side unit jacks are connected to the secondaries of the left hand side unit. The current terminals of the right-hand side jacks are connected to the right hand side transformer primaries; the secondaries of that unit are brought to the current terminals of the left-hand side unit.

In the "M" position of the connection switch, the current terminals of the jacks of both the units are connected to their respective primaries. But the secondaries of one unit are connected to the voltage terminals of the jacks of the other unit as shown in Fig. 6(B).

NETWORK SIMULATION BY MEANS OF TAC UNITS

When the units of TAC are interconnected to represent a network, there are four separate circuits formed, the in-phase and quadrature-phase circuits of voltage and current. These circuits are electrically separate but are linked by the magnetic flux in the transformers. When the analyzer transformers representing an element are interconnected to the transformers representing another, the interconnections between the current and voltage circuits in the two units must be such as to satisfy the relationship between the voltages and currents in the actual elements. Also, the analyzer interconnections have to satisfy the network laws at all nodes and round all meshes, viz., the algebraic sum of all the currents at a node must be zero and the algebraic sum of all the voltages around the mesh must be zero.

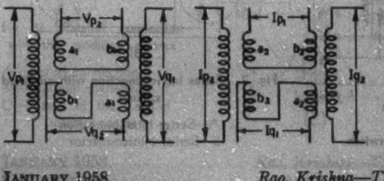


Fig. 2 (left). Transformer representation

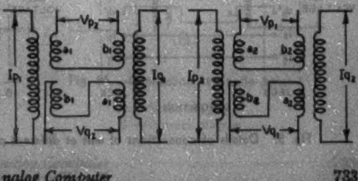


Fig. 3 (right). Mutual impedance representation

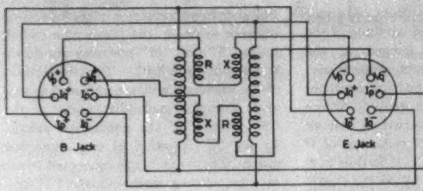


Fig. 4 (above). Connections of standard impedance unit

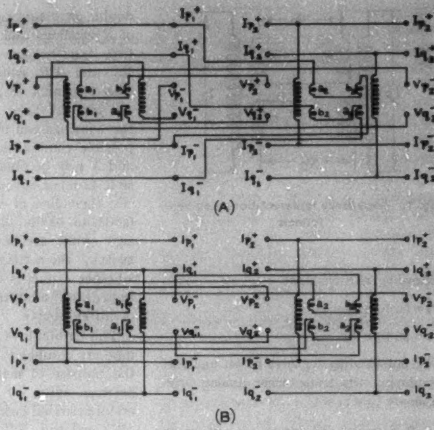


Fig. 6 (right). Interconnections of the universal unit

A—Operated as a transformer

B—Operated as a mutual impedance

The interconnections between the units of TAC are made at a plugging board by means of series and parallel interconnectors. The series connection between units is made by means of the series connectors whereas connections at a node point are made by means of parallel interconnectors.

THE SERIES CONNECTION

The method of connecting two units Z_1 and Z_2 in series by means of the series interconnector is shown in Fig. 7(A) and 7(B). The series interconnector consists of two sockets the corresponding terminals of which are connected to each other. The units Z_1 and Z_2 are put in series by inserting the "E" jack of Z_1 and the "B" jack of Z_2 into the sockets of the series interconnector. From an inspection of the "B" and "E" jacks it will be noticed that as far as the current coils are concerned, the terminals are connected in parallel whereas the potential coils are put in series. This ensures that the currents in the two impedances are the same but the voltages add up between the terminals formed after interconnection of the two impedances. The "B" jack of Z_1 and the "E" jack of Z_2 then represent the two ends of a unit equivalent to Z_1 and Z_2 in series. The interconnection between the coils of the units are shown in Fig. 8.

PARALLEL CONNECTION

The parallel connection of two units is made by means of two parallel interconnectors at the two ends of the units. The method of connecting two impedances Z_1 and Z_2 in parallel are shown in Fig. 9.

The parallel interconnector consists of three sockets. The ends of the units are brought to the outer sockets; at the central socket the connected end is available for further interconnection. The internal connections between the sockets of a parallel interconnector are shown in figure 9(C). When the "B" jack of Z_1 is inserted into socket 1 and "B" jack of Z_2 is inserted into socket 3, the various terminals of the two units appearing at the socket points are shown. It is seen that the internal connections between sockets 1 and 3 are such that the beginning of the voltage coils of the two units are connected together, while the ends of the current coils of Z_1 are connected to the beginning of the current coils of Z_2. The middle socket is connected to the outer sockets 1 and 3 in such a manner that the socket 2 has the beginning of the current coils of Z_1, the beginnings of the voltage coils of Z_1 and Z_2 tied up and the ends of the current coils of Z_2. Thus the socket 2 fully represents the beginning of the unit equivalent to Z_1 and Z_2 in paral-

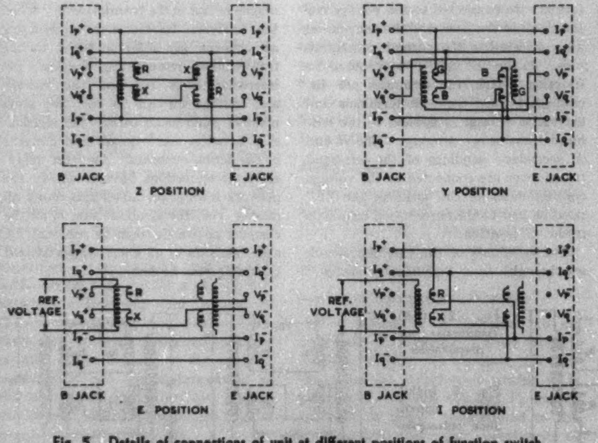


Fig. 5. Details of connections of unit at different positions of function switch

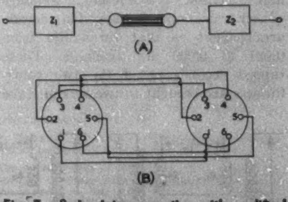


Fig. 7. Series interconnection with modified scheme

A—Series interconnection

B—Series interconnector

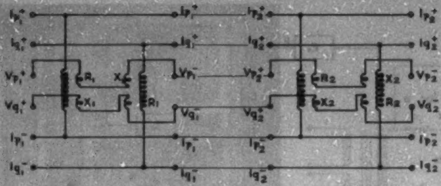

$I_{p_1}^+$
$I_{q_1}^+$
$V_{p_1}^+$
$V_{q_1}^+$
$I_{p_1}^-$
$I_{q_1}^-$
R_1
X_1
$V_{p_1}^-$
$V_{q_1}^-$
$I_{p_2}^+$
$I_{q_2}^+$
$V_{p_2}^+$
$V_{q_2}^+$
$I_{p_2}^-$
$I_{q_2}^-$
R_2
X_2
$V_{p_2}^-$
$V_{q_2}^-$

Fig. 8 (above). Series connection of two impedance units

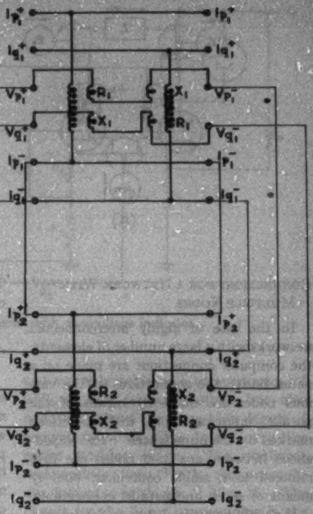

$I_{p_1}^+$
$I_{q_1}^+$
$V_{p_1}^+$
$V_{q_1}^+$
$I_{p_1}^-$
$I_{q_1}^-$
R_1
X_1
$V_{p_1}^-$
$V_{q_1}^-$
$I_{p_2}^+$
$I_{q_2}^+$
$V_{p_2}^+$
$V_{q_2}^+$
$I_{p_2}^-$
$I_{q_2}^-$
R_2
X_2
$V_{p_2}^-$
$V_{q_2}^-$

Fig. 10 (right). Parallel connection of two impedance units

lel. The middle socket of a similar interconnector to which the "E" jacks of Z_1 and Z_2 are brought represents the end of the unit equivalent to Z_1 and Z_2 in parallel. The actual interconnection made between the units Z_1 and Z_2 is shown in Fig. 10.

CONNECTIONS AT A NODE POINT

The computer connections at a node point are made by means of parallel interconnectors. The connections are made in a cyclic order with convenient directions assumed for each impedance. At each node in the network, the current coils of the units coming to that node form a closed loop thus satisfying the network node law $\Sigma I = 0$.

Fig. 11 shows the computer connection for the node N formed of elements Z_A, Z_B, Z_C, Z_D, and Z_E. The directions for the impedances meeting at a node are so assumed as to form two groups of elements one with those whose directions point towards the node and the other with those whose directions point away from the node; also similar directions appear successively and there are no alternate reversals of direction. When making the nodal connection, the "E" jacks are all connected to a set of parallel interconnectors and the "B" jacks to another set. Finally the two sets are connected in series by a 6-core cord connecting the two sockets left over. It will be noticed that at a parallel connector at which the end jacks are connected, the sequence for the units is from bottom to top, whereas at an interconnector where the beginning jacks are connected, the sequence is from top to bottom.

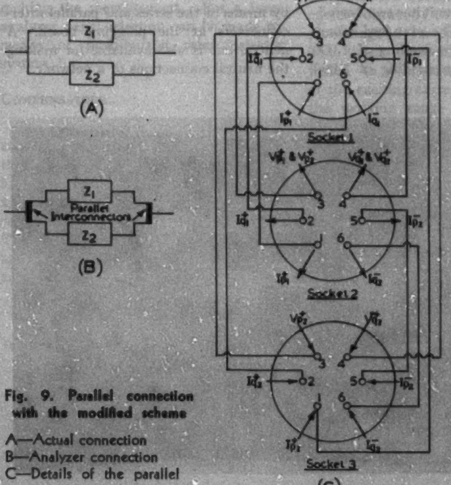

Z_1
Z_2
(A)
Z_1
Parallel Interconnectors
Z_2
(B)
$V_{p_1}^+$
$V_{q_1}^+$
3
4
$I_{q_1}^+$
2
5
$I_{p_1}^-$
1
6
$I_{p_1}^+$
$I_{q_1}^-$
Socket 1
$V_{p_1}^+$ & $V_{p_2}^+$
$V_{q_1}^+$ & $V_{q_2}^+$
3
4
$I_{q_1}^-$
2
5
$I_{p_2}^-$
1
6
$I_{p_1}^+$
$I_{q_2}^-$
Socket 2
$V_{p_2}^+$
$V_{q_1}^+$
3
4
$I_{q_2}^+$
2
5
$I_{p_2}^-$
1
6
$I_{p_1}^+$
$I_{q_2}^-$
Socket 3
(C)

Fig. 9. Parallel connection with the modified scheme

A—Actual connection
B—Analyzer connection
C—Details of the parallel interconnector

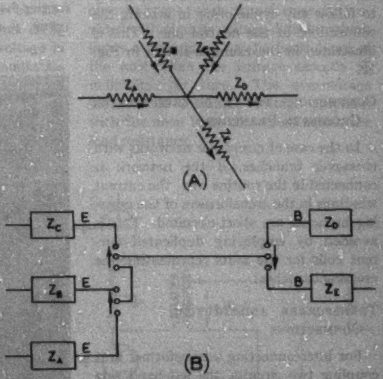

Z_B
Z_C
Z_A
Z_D
Z_E
(A)
Z_C
E
Z_B
E
Z_A
E
B
Z_D
B
Z_E
(B)

Fig. 11. Connection at a node point

A—Actual diagram
B—Analyzer diagram

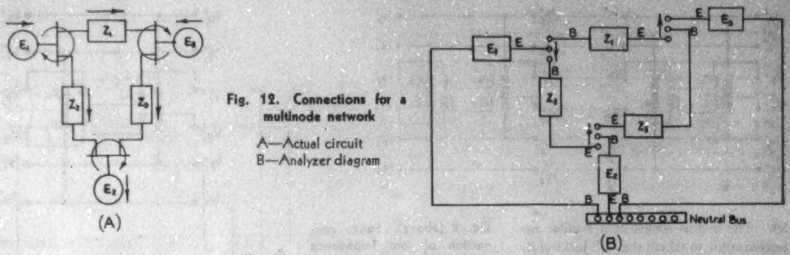


Fig. 12. Connections for a multinode network

A—Actual circuit
B—Analyzer diagram

Connections for a Network With Multiple Nodes

In the case of highly interconnected networks with a large number of elements, the computer connections are made on a nodal basis; the connections at the various nodes are made by means of the parallel interconnectors, with directions marked on the impedances. The impedances between any two nodes are first reduced to a single equivalent one by means of series and parallel connections.

It should be noticed that as the current coil of a branch unit is common to the two nodes to which that unit is connected, the current coils circuit for the network is complete even when one nodal connection is not made. Advantage is taken of this fact for the case of power system networks where the earth terminal forms a node of several branches. Hence for making the earth connections, a neutral bus of sockets may be employed at which only the voltage coils are connected to each other and current coil connections are not made at all. It is not necessary to follow any cyclic order in making the connections at the neutral bus. This is illustrated in the example shown in Fig. 12.

Connections for a Network With Crossover Branches

In the case of nonplanar networks with crossover branches, if the network is connected in the routine way, the current windings in the transformers of the crossing branches are short-circuited. This is avoided by employing duplicated current coils for the units representing the crossing branches.

Transformer and Mutual Connections

For interconnecting a transformer unit coupling two circuits, the left-hand side of a universal unit is assumed to be a load in the first circuit and the right-hand side unit is assumed to be the source supplying the second network. The connections for a mutual unit in a circuit are the same as for the transformer unit.

Constructional Features of TAC

TAC consists of an assembly of standard universal units mounted in racks, a measuring equipment at which the analyzer answers are read and a central plugging board at which the interconnections between the units are made. The other auxiliaries of the computer are the power supplies for the filament and plates of the compensating tubes and a source of 50-cycle stabilized excitation. The front and side views of the computer developed and built in this laboratory are shown in Figs. 13 and 14 respectively.

Each universal unit is inserted into the rack from behind and the units can be drawn out at the back for inspection or repair. Each universal unit is composed of two ordinary units linked by the connection switch at the center of the front panel. The ordinary unit has two 3-winding transformers. The standard transformer has a primary winding of 1,000 turns, a compensator winding of 1,000 turns with a center tap and two secondaries of 999 turns, tapped at ones, tens, and hundreds of turns. The two secondaries are wound one above the other so as to equalize the leakage between the primary and the secondaries. The four mumetal transformers of a universal unit are mounted on a chassis which is bolted to an angle frame supporting the front and back panels. Behind the transformers on the chassis are the compensating tubes. On the front panel are the sign switch, the function switch, the *RG* and *XB* decade switches of the two units and the *STM* connection switch interconnecting two units. Provision is made at the back panel for the bringing of the *HT* and filament supplies to the compensator circuits.

At the centralized plugging board at which the analyser interconnections are made, the two terminals of each unit are available as its beginning and end jacks. The analyzer interconnections are made by means of the series and parallel interconnectors at the plugging board. A neutral bus is also available for making the neutral connections of a network.

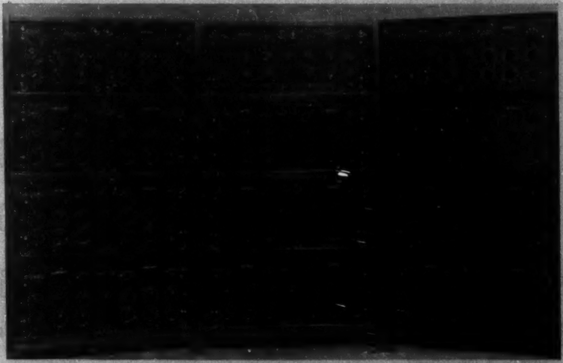


Fig. 13. Front view of TAC

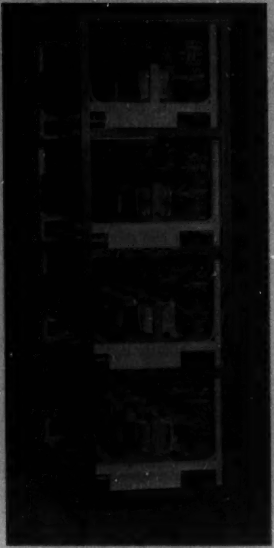


Fig. 14. Side view of TAC

For TAC, the panel selection for metering is done by means of selector switches at the metering desk. At the metering desk, it is possible to use either the direct system or the null method for making the measurement. For a large analyzer panel selection can be made by means of push-button relay selector system.

COMPENSATORS

The accuracy of a transformer analog computer depends to a great extent on the accuracy of the individual transformers. TAC transformers are designed to have a small ratio of leakage to mutual reactance by using mumetal cores. The exciting currents of the transformers are supplied by means of auxiliary electronic compensators.

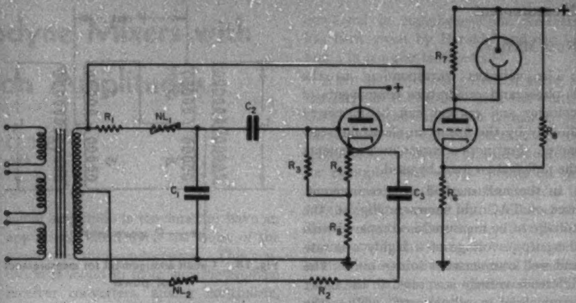


Fig. 15. Nonlinear compensator with overvoltage indicator

The linear compensator developed for TAC is similar to that suggested by Humphrey Davies and Slemon.[4] On the core of the transformer whose exciting current has to be compensated is wound a compensator winding of the same number of turns as the primary but provided with a center tap. The voltage developed across the full winding is applied to a resistance-capacitance phase shifting network. The voltage across the capacitor is applied to the input terminals of the cathode follower. The output voltage of the cathode follower is applied between the ground and center tap of the compensation winding and this forces the exciting current for the transformer through the bottom half of the compensation winding, thus reducing the current drawn by the primary to a small percentage of its uncompensated value. The linear compensator could be adjusted to compensate for about 97% of the exciting current at any voltage in the working range but the compensation at other voltages in the range drops down to about 90%.

The improved nonlinear compnseator developed here overcomes this defect and not only achieves a compensation of 97% at any primary voltage but also maintains about the same high degree of compensation at all voltages over the entire working range. The nonlinear compensator circuit with the overvoltage indicator is shown in Fig. 15. The circuit for the linear compensator is the same as for the nonlinear compensator with the exception of the nonlinear resistors NL_1 and NL_2. In the nonlinear compensator a combination of linear and nonlinear resistors are employed but the two nonlinear resistors have nonlinearities of opposite character. NL_1 is a nonlinear resistor whose resistance decreases as the voltage across it increases whereas NL_2 is a nonlinear resistor whose resistance increases as the voltage across it increases.

In order to avoid possible errors due to over voltages, each transformer is provided with an over voltage detector which lights a neon lamp on the back panel of the unit when the voltage exceeds 15 millivolts per turn. The overvoltage indicator is a simple voltage amplifier with the neon lamp connected across the load resistance.

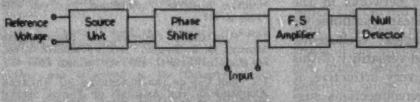


Fig. 16. Null method (TAC)

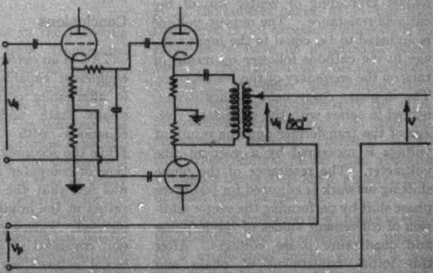


Fig. 17 (right). Direct method of measurement

Measurement

Null Method

The voltages corresponding to the in-phase and quadrature components of voltages and current can be measured directly by the use of an electronic voltmeter. For more accurate measurements the null system must be used.

In the null method of measurement, used on TAC and shown in Fig. 16, the voltage to be measured is compared with the output voltage of a highly accurate and well compensated source unit. The difference voltage is applied to the input terminals of a null detector. When the null detector reads zero, the setting of the source unit gives the required quantity. In order to overcome the effects of any phase-angle shift in the analyzer, provision is made for shifting the phase of the reference voltage through a small angle, keeping the amplitude unaltered. A frequency selective amplifier is interposed before the null detector to suppress the undesirable harmonics, that might be present. The null detector used is a vacuum tube voltmeter and the frequency selective amplifier is of the feedback type with a twin T-rejector filter network interposed between two cathode followers.

Direct Method

In order to facilitate rapid measurements, a direct system of measurement of the quantities in polar form and the active and reactive powers is necessary. The direct system of measurement used on the TAC is illustrated in Fig. 17. The quadrature component voltage is applied to a resistance capacitance network through a linear cathode follower. The voltage between the center tap of the cathode load resistor and the junction of resistance and capacitance, is applied to the primary of a standard transformer through the interposition of cathode followers. The arrangement illustrated avoids grounding of center tap of the cathode resistance. The output voltage is adjusted to be equal to the input voltage to the cathode follower by means of taps on the secondary of the transformer. The in-phase component voltage is connected in series with the secondary voltage of the transformer and the resultant voltage is measured by a vacuum-tube voltmeter. The resistance of the phase shifting network is adjusted for 90-degree phase shift by performing the vector addition of convenient values of the in-phase and quadrature phase voltages. After this initial adjustment of phase and amplitude the equipment measures the vector value of the voltage across any unit. The current is also measured in a similar way.

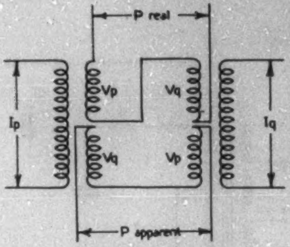


Fig. 18. Circuit arrangement for measurement of power

For measurement of power the arrangement shown in Fig. 18 is employed. Here voltages representing the in-phase and quadrature phase of currents are applied to the primaries and the decade switches of the ganged secondaries are set at per unit values of the in-phase and quadrature phase voltages representing the voltage. The active and reactive powers are then measured as voltages.

The circuit arrangement of Fig. 18 satisfies the following equations:

$$\text{Power} = V^*I = (V_p - jV_q)(I_p + jI_q)$$

$$(V_pI_p + V_qI_q) + j(V_pI - V_qI_p)$$

$$\text{The active power} = V_pI_p + V_qI_q$$

$$\text{Apparent power} = V_pI_q - V_qI_p$$

In the actual equipment built here, for direct system of measurement, a selector switch is used for selecting I_p, I_q, I, V_p, V_q, V, P real, and P apparent. The metering connections are made with the aid of cascaded panel selector switches by means of which it is possible to make measurements of voltage either across a single unit or from one unit to any other selected point in the system.

Conclusions

TAC is an extremely useful computing equipment capable of solving problems in alternating current networks in the steady state, and linear simultaneous equations with complex coefficients. While the accuracy and versatility of this type of computer are very attractive, it was felt that the interconnections suggested in the original scheme were complicated. This was largely because each unit consisted of four terminals for external connections instead of two terminals as in an ordinary A-C Network Analyzer.

By the introduction of the special parallel interconnectors it has now been made possible to utilize, in the large majority of cases, the actual circuit diagram itself for setting up problems on the analyzer. The interconnections for the series and parallel connections are easy to make and follow. Furthermore, with the double-ended units, it is possible to make the analyzer connections on a centralized plugging board and thus reduce the time and labor involved in solving problems.

A separate paper dealing with the application of TAC for the solution of problems in a-c machinery, servomechanisms, and other related problems of engineering industry is contemplated.

TAC would be very useful to power system engineers and electric machine designers. It should also prove itself to be an invaluable tool for research and development work in the field of automatic control engineering.

References

1. Improvements in and Relating to Calculating Machines, C. L. Blackburn. *British Patent no. 476,164*, 1937.

2. Improvements in and Relating to Calculating Machines, C. L. Blackburn. *British Patent no. 476,165*, 1937.

3. Application of the Transformer Analogue Computer, J. R. Barker. *British Journal of Applied Physics*, London, England, vol. 7, 1956, p. 303.

4. Transformer Analogue Network Analysers, M. W. Humphrey Davies, G. R. Slemon. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 100, pt. II, 1953, pp. 469–86.

5. An Electrical Calculating Machine, R. R. M. Mallock. *Proceedings*, Royal Society, London, England, vol. 140, series A, 1933, p. 457.

6. Improvements in and Relating to Calculating Machines, R. R. M. Mallock. *British Patent no. 389,524*, 1933.

7. Network Analysis. *Electrical Times*, London, England, vol. 117, pp. 307–11.

8. Some Experiences with a British A. C. Network Analyser, G. Lyon. *Proceedings*, Institution of Electrical Engineers, vol. 97, pt. II, 1950, p. 697.

9. Equivalent Circuits of Electric Machinery (book), G. Kron. John Wiley & Sons, Inc., New York, N. Y., 1951.

10. Development of the Transformer Analogue Analyser and Its Application to Electrical Machines, M. K. E. Ishmail. *Ph.D. Thesis*, University of London, London, England, 1955.

11. Linear Algebraic Computation by Multiwinding Transformers, F. L. Ryder. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 259, 1955, p. 427.

12. Equivalent Circuits for Single-Phase Motors, G. R. Slemon. *AIEE Transactions*, vol. 74, pt. III, 1955 (Feb. 1956 section), pp. 1335–43.

13. The Use of the Blackburn AC Network Analyser in the Analysis of Power System Faults, J. H. Banks, K. C. Parton. *Proceedings*, Institution of Electrical Engineers, pt. C, 1956, p. 346.

14. An Improved Compensator for Transformers, P. Venkata Rao. *U. S. Patent no. 644,685*, *Indian Patent no. 56,877*, 1957.

Small-Signal Heterodyne Mixers with Excessive Injection Amplitudes

JACK F. CLINE
MEMBER AIEE

Synopsis: The theory of small-signal heterodyne mixers operating with very large injection potentials is described. The mixer properties investigated are the conversion transconductance, the relative amplitudes of the difference-frequency and injection frequency components in the output, and the amplitude of the injection modulation-frequency component in the output in the case where the injection potential is amplitude-modulated to a small degree, as by power-supply ripple or noise, for example. The theoretical analysis is based upon 2- and 3-segment piecewise-linear approximations of the current and incremental transfer conductance curves, so that the mixer can be treated as a switch-type or commutator-type modulator. The method of constructing the approximations from the original curves is described. The result is a set of universal performance curves for mixers of the particular class chosen to illustrate the method. The method can be extended to other classes of mixers.

THIS REPORT describes the operation of small-signal heterodyne mixers in which the injection potential is a biased sinusoid which may be permitted to become much larger than in ordinary receiver or spectrum analyzer applications. Such excessive injection potentials are found either in applications where it is desired to obtain a decreasing conversion transconductance with an increasing injection potential or in applications where, for reasons independent of mixer performance considerations, it is either necessary or expedient to tolerate an excessive injection potential.

The term "small signal" is used here to indicate the usual condition in which signal amplitude is too small to have an appreciable effect upon the value of the conversion transconductance. This condition is found in most superheterodyne receiver converters, analog multipliers, spectrum analyzers, etc. The mixer properties which are investigated here are the conversion transconductance, the relative amplitudes of the difference-frequency and injection-frequency components in the output, and the amplitude of the injection modulation-frequency component in the output in cases where the injection potential is amplitude modulated to a small degree, as, for example, by noise, or power-supply ripple.

Method of Analysis

The method of analysis described here is illustrated by its application to a class of heterodyne mixers having characteristic curves of the particular shape shown as solid-line curves in Fig. 1. This particular class of heterodyne mixers is one for which there was a special application in the sponsored research on which this report is based, but the method can be extended easily to other classes of mixers. The essential properties of the solid curves in Fig. 1 are that the output current characteristic i has two horizontal sections at different levels while the incremental transfer conductance characteristic g has two horizontal sections both at zero level. Both curves are plotted with the instantaneous injection potential e_i as the absissa. The details of the curved transitional sections which join the horizontal sections are unimportant.

It will be convenient in the following discussion to use the term "transitional potential range" in referring to the range of instantaneous injection potential e_i over which the curved sections of i and g extend. In most mixer applications the swing of the injection potential is usually not much larger than the transitional potential range. This is true in many superheterodyne radio receivers, for example. A discussion of the conversion transconductance of small-signal heterodyne mixers which operate with injection potential swings of the magnitude often employed in superheterodyne receivers has been given by Herold,[1] and can be found in popular textbooks.[2,3]

When the injection potential swing is not large in comparison with the transitional potential range, the details of the curve shape are quite important in the quantitative analysis. Herold's analysis, for example, uses several points on the transitional part of the curve in order to take adequate account of the details of the shape. On the other hand, when the injection potential swing is very large in comparison with the transitional potential range, as in the case discussed in this paper, the fine details of the curved section are found to be relatively unimportant. This makes it possible to replace the solid curves in Fig. 1 with the piecewise-linear curves i' and g', which are drawn as dashed lines in the same figure. The mixer then becomes essentially a switch-type or commutator-type circuit, which has been discussed from a modulation standpoint by Caruthers[4] and by Peterson and Hussey.[5] For any particular mixer, the curves corresponding to the solid lines of Fig. 1 must be obtained experimentally and it is difficult in general to obtain suitable mathematical expressions for them. The piecewise-linear curves, on the other hand, are easily described in mathematical terms. The problem is to find a simple way of constructing the piecewise-linear curves so that when they are analyzed mathematically, the results will be, as nearly as possible, the same as the results obtained from the solid curves. This is one of the problems discussed in this paper.

Consider a mixer in which there is a resistive load R_L and an applied signal potential of instantaneous value e_s. The incremental transfer conductance g is then defined as the partial derivative of the instantaneous load current i with respect to e_s. A distinction must sometimes be made between this incremental transfer conductance g of the circuit as a whole and the familiar incremental transconductance g_m of the control device alone, but this distinction is not necessary when R_L is so small as to have no appreciable effect upon i. Even under circumstances where g and g_m do differ appreciably, they usually vary in approximately the same proportional as functions of the instantaneous injection potential e_i, so that when relative rather than absolute values are being considered, g and g_m may be used interchangeably.

The above definition of g applies to both single-input and double-input mixers. In a single-input mixer, e_s and e_i are applied in series to the same input termi-

---

Paper 57-1027, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7-11, 1957. Manuscript submitted December 26, 1956; made available for printing August 12, 1957.

JACK F. CLINE is with the Stanford Research Institute, Menlo Park, Calif.

The work reported here was carried out while the author was employed in the Department of Electrical Engineering and the Engineering Research Institute of the University of Michigan, Ann Arbor, Mich. The work was sponsored under contract no. *DA-36-039 sc-623203*, Signal Corps, Department of the Army, under Department of the Army project no. *3-99-04-042*. The author is indebted to the staff of the Stanford Research Institute for assistance in preparing the manuscript for publication.

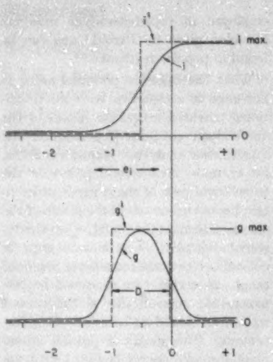
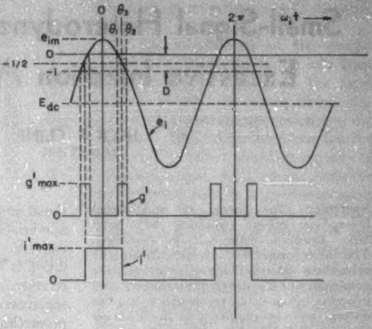


Fig. 1 (left). Output current (i or i') and incremental transfer conductance (g or g') as functions of the instantaneous total injection potential e_i; the dashed curves i' and g' are the piecewise-linear approximations

Fig. 2 (right). Relation between instantaneous values of injection potential e_i, transfer conductance g', and output current i' as functions of time when the piecewise-linear approximations are used



nal-pair, so that g is equal to the partial derivative of i with respect to either e_s or e_i. In a double-input mixer, however, e_s and e_i are applied to separate terminal-pairs and in general they have unequal effects upon i. Therefore g is, in general, not equal to the partial derivative of i with respect to e_i in a double-input mixer. Upon causal inspection of Fig. 1, it might be concluded from the shapes of the solid curves of i and g versus e_i that a derivative relationship exists and that the solid curves of Fig. 1 therefore represent a single-input mixer. Even if this were true, however, the piecewise-linear curves i' and g' certainly do not possess a derivative relationship. The method used in the analysis which follows is such that a derivative relationship between the i' and g' curves is not required; consequently, the method may be applied to either single-input or double-input mixers.

In constructing the piecewise-linear curves, the discontinuities in g', which define the dynamic range D, are placed so that the area under g' is equal to the area under g and also so that the centers of gravity of the two areas are located at the same value of e_i. For convenience, the origin of the e_i scale is placed at the right-hand limit of D, while the e_i scale units are chosen so that D equals unity. This serves to normalize the injection potential with respect to the dynamic range and thereby simplifies some of the mathematical expressions. The discontinuity in i' is placed so that the area under i' is equal to the area under i, as a result of which the discontinuity falls somewhere near $e_i = 1/2$.

It will develop subsequently that the three mixer properties with which this paper is concerned can be expressed in terms of the zero- and first-order coefficients of i and g when they are expanded in Fourier series as functions of time. The reason for drawing the piecewise-linear curves with the equal area and center of gravity properties described above is that their zero- and first-order coefficients are approximately the same as those of the solid curves when this is done. The reason for this can be explained in terms of the g and g' curves as an example. Consider the transitional range of injection potential for which g has an appreciable value. This will be an interval of e_i somewhat larger than D, but much smaller than the total swing of the injection potential. If this total swing extends quite far beyond both limits of the transitional range, time, and potential will be quite linearly related within this range, and consequently the areas under g and g', which have been made equal to each other when these quantities are plotted against e_i, also will be approximately equal to each other when g and g' are plotted against time. Now let g and g' be expanded in Fourier series as functions of time, with the zero reference chosen so that they are even functions. If t is time in seconds and if ω_i is the injection frequency in radians per second, the Fourier coefficient of order 1 is

$$a_1 = (2/\pi)\int_0^\pi g'' \cos(\omega_i t)\, d(\omega_i t) \tag{1}$$

where g'' represents either g or g'. Now if θ_1 and θ_2 are the angular limits of $\omega_i t$ corresponding to the interval of potential or time being considered, and if this interval is quite small, as has been assumed, the variation of cos $(\omega_i t)$ over this interval is small enough so that $\omega_i t$ may be treated simply as a constant β, where β is some angle between θ_1 and θ_2. Now equation 1 can be rewritten as an approximation

$$a_1 = \frac{2\cos\beta}{\pi}\int_{\theta_1}^{\theta_2} g''\, d(\omega_i t) \tag{2}$$

The integral in equation 2 is simply the area under either g or g' when plotted against time, and since these areas have been made approximately equal, we conclude that this integral has approximately the same value for g' as for g. Its value can be obtained from the g' curve by choosing the angular limits θ_1 and θ_2 to be the exact limits of the nonzero segment of g', as indicated in Fig. 2. The value of β is the same for both g' and g because of the equal center of gravity condition. It follows that a_1 is the same for both g' and g. A similar analysis justifies the equal area conditions for i' and i.

Conversion Transconductance

It has been noted previously that when the load resistance R is not small a distinction may be made between the incremental transfer conductance g (or g') of the circuit as a whole, acting as a straight amplifier, and the incremental transconductance g_m of the tube, or other control device, alone. Similarly, for a heterodyne mixer, a distinction may be made between what might be called the conversion transfer conductance of the circuit and the conversion transconductance of the control device alone. However, for simplicity, this distinction will be avoided here, and the symbol G_c will be used to represent either or both quantities.

Consider a heterodyne mixer circuit whose characteristics are represented by by g' curve in Fig. 1. The development of the conversion transconductance G_c follows the familiar pattern, with a few exceptions. For the reader's convenience, we review it briefly here. Suppose that a signal potential e_s is given as

$$e_s = E_s \cos(\omega_s t) \tag{3}$$

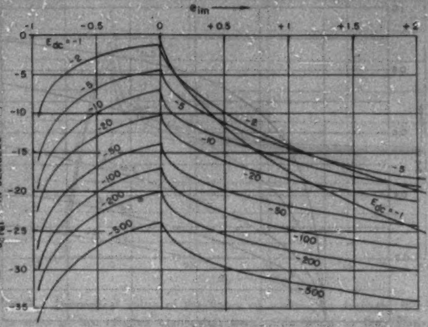


Fig. 3. Theoretical values of relative conversion transconductance G_{crel} obtained from equations 13 and 14, expressed in decibels; the symbol E_{dc} represents the d-c injection bias potential and e_{im} represents the maximum instantaneous total injection potential

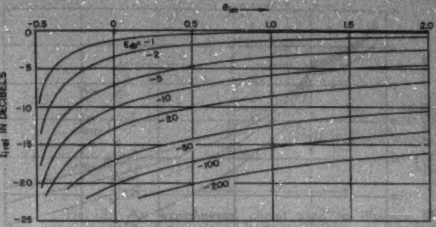


Fig. 4. The relative amplitude I_{irel} of the injection frequency component in the output current, according to equation 19, expressed in decibels

where E_s is very small. Let this potential be applied to the appropriate mixer terminal-pair. Since the transfer conductance g' was defined previously as the partial derivative of the output current with respect to e_s, we have

$$i_o=\int g'de_s \tag{4}$$

where i_o is the output current (not shown in Fig. 1) corresponding to g'. Since E_s is very small, g' is essentially independent of e_s, so that

$$i_o=g'\int de_s=C+g'e_s \tag{5}$$

where C is the constant of integration. Now let the injection voltage be an even function

$$e_i=E_{dc}+E_i\cos(\omega_i t) \tag{6}$$

Then g' and e_i are related in time as shown in Fig. 2, and g' can be written as a Fourier cosine series

$$g'=b_0/2+b_1\cos(\omega_i t)+b_2\cos(2\omega_i t)+\text{etc} \tag{7}$$

If equations 3 and 7 are now substituted for g' in 5, the current becomes

$$i_o=C+(b_0/2)E_s\cos(\omega_s t)+(b_1/2)E_s\cos(\omega_s-\omega_i)t+(b_1/2)E_s\cos(\omega_s+\omega_i)t+(b_2/2)E_s\cos(\omega_s-2\omega_i)t+(b_2/2)E_s\cos(\omega_s+2\omega_i)t+\text{etc.} \tag{8}$$

One of the terms in equation 8 has an angular frequency $(\omega_s-\omega_i)$; the ratio of the coefficient of this term to the amplitude E_s of the signal potential is defined as the conversion transconductance G_c. Evidently,

$$G_c=b_1/2 \tag{9}$$

which is the well-known relationship. The value of b_1 for the piecewise-linear function g' is given by the formula for the Fourier coefficient of an even function:

$$b_1=(2/\pi)\int_0^\pi g'\cos(\omega_i t)d(\omega_i t) \tag{10}$$

From either Fig. 1 or Fig. 2, it is seen that g' is equal to a constant maximum value g'_{max} over the dynamic interval and equal to zero elsewhere, so that if θ_1 and θ_2 are the angular limits of the g' pulse, as indicated in Fig. 2,

$$b_1=(2/\pi)\int_{\theta_1}^{\theta_2} g'_{max}\cos(\omega_i t)d(\omega_i t) \tag{11}$$

By substituting equation 11 in place of b_1 in equation 9 and performing the indicated integration, we obtain the following expression for the conversion transconductance:

$$G_c=(g'_{max}/\pi)(\sin\theta_2-\sin\theta_1) \tag{12}$$

The difference $(\sin\theta_2-\sin\theta_1)$ can vary from zero to unity, so that it can serve conveniently as a definition of the relative conversion transconductance, which will be denoted here by the symbol G_{crel}. Since θ_1 and θ_2 are quantities not easily measured, it is more useful to have G_{crel} expressed in terms of more easily measured quantities such as certain components of the injection potential. Three component values suggest themselves as possibilities: the d-c or bias component E_{dc}, the peak value E_{im} of the a-c component, and the peak instantaneous value e_{im} of the total injection potential (see Fig. 2). If two of these are known, the third is determined, so that only two are needed in the expression for G_{crel}. It happens that G_{crel} is most sensitive to changes in e_{im}, so that it seems logical to use e_{im} as one of the values in the new expression. E_{im} and E_{dc} are in general rather large quantities, while e_{im} is the small difference between them, so that if E_{im} and E_{dc} are taken as the measured values, a small fractional error in either results in a large fractional error in e_{im}. Consequently it is desirable to use e_{im} together with either one of the other two in the new expression for G_{crel}. E_{dc} has been chosen here. From trigonometry, we find that

$$\sin\theta_2-\sin\theta_1=G_{crel}=\sqrt{1-\left(\frac{-E_{dc}-1}{-E_{dc}+e_{im}}\right)^2}-\sqrt{1-\left(\frac{-E_{dc}}{-E_{dc}+e_{im}}\right)^2} \tag{13}$$

The last term does not apply when e_{im} lies between zero and minus one, so that

$$\sin\theta_2-\sin\theta_1=G_{crel}=\sqrt{1-\left(\frac{-E_{dc}-1}{-E_{dc}+e_{im}}\right)^2} \tag{14}$$

Values of G_{crel}, as given by equations 13 and 14 and expressed in decibels, are plotted in Fig. 3 for the more interesting ranges of values.

It is interesting to note that when $(-E_{dc})$ is large in comparison with both e_{im} and unity, the relative conversion transconductance becomes approximately equal to $\sqrt{2/(-E_{dc})}$, which suggests that the circuit might be useful as an analog for an inverse square-root function. In any practical circuit, where the transconductance is more accurately described by g than by g' (Fig. 1), it would be necessary to hold e_{im} constant at zero or some small value in order to achieve this result.

Injection Frequency Component

Consider again the hypothetical mixer whose characteristics are represented by the curves in Fig. 1. Let the piecewise-linear approximation i' replace i in the

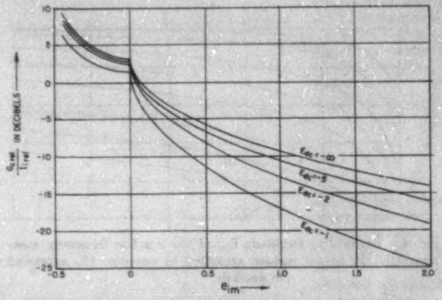


Fig. 5. The ratio of the relative conversion transconductance of Fig. 3 to the relative current amplitude of Fig. 4, in decibels; the limiting case where E_{dc} is infinite is given by equations 20 and 21

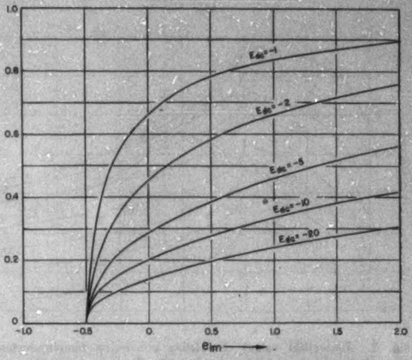


Fig. 6. The relative value I_{dcrel} of the d-c component of the output current, according to equation 24; all curves approach unity as e_{im} becomes very large

analysis. (Note that i' is not the same as the i_o of the preceding section, which was obtained by integrating g' with respect to e_s.) As shown in Fig. 2, i' is an even function of time, so that it can be represented by a Fourier cosine series:

$$i'=B_0/2+B_1\cos(\omega_i t)+B_2\cos(2\omega_i t)+\text{etc.} \tag{15}$$

The coefficient B_1 is evidently the amplitude of the injection-frequency component of the output current, and consequently will be denoted as I_i in the remainder of the discussion. From the formula for the Fourier coefficient of an even function, we have

$$I_i=(2/\pi)\int_0^{\pi} i'\cos(\omega_i t)d(\omega_i t) \tag{16}$$

Let θ_3 denote the value of $\omega_i t$ for which $e_i=-1/2$. It is evident from Fig. 2 that i' has a constant maximum value $i'_{\max}$ between 0 and θ_3 and a value of zero between θ_3 and π, so that equation 16 may be rewritten as

$$I_i=(2/\pi)i'_{\max}\int_0^{\theta_3}\cos(\omega_i t)d(\omega_i t)=(2/\pi)i'_{\max}\sin\theta_3 \tag{17}$$

As a function of θ_3, I_i evidently has a maximum value equal to $(2/\pi)i'_{\max}$ when $\theta_3=\pi/2$ and a relative value

$$I_{irel}=\sin\theta_3 \tag{18}$$

From trigonometry, I_{irel} can be evaluated in terms of E_{dc} and e_{im} as

$$I_{irel}=\sqrt{1-\left(\frac{-E_{dc}-1/2}{-E_{dc}+e_{im}}\right)^2} \tag{19}$$

Values of equation 19, expressed in decibels, are plotted in Fig. 4 as functions of e_{im}, with E_{dc} as a parameter.

By dividing either equations 13 or 14 by 19, depending upon the value of e_{im}, we obtain a number which is proportional to the ratio of the amplitude of the desired difference-frequency component to the amplitude of the undesired injection-frequency component in the output current. Since this number is only proportional to, not equal to, the ratio of the amplitudes, its actual value is of less interest than the way in which its value changes with e_{im} for different values of E_{dc}. This is shown in Fig. 5, where ordinates are expressed in decidels. The curves of Fig. 5 may be obtained simply by subtracting the curves of Fig. 4 from those of Fig. 3, for the corresponding values of E_{dc}. In the limiting case where E_{dc} approaches minus infinity,

$$\frac{G_{crel}}{I_{irel}}=\frac{\sin\theta_2-\sin\theta_1}{\sin\theta_3}=\sqrt{\frac{2(e_{im}+1)}{2e_{im}+1}}-\sqrt{\frac{2e_{im}}{2e_{im}+1}} \tag{20}$$

for the case where e_{im} is positive and

$$\frac{G_{crel}}{I_{irel}}=\frac{\sin\theta_2-\sin\theta_1}{\sin\theta_3}=\sqrt{\frac{2(e_{im}+1)}{2e_{im}+1}} \tag{21}$$

for the case where e_{im} lies between $-1/2$ and zero. These are relative values. The actual value of the ratio of the amplitude of the difference-frequency component of the output current to the amplitude of the injection-frequency component of the same current can be obtained, in decibels, by adding the term $20\log_{10}(E_s g'_{\max}/2i'_{\max})$ to the curves of Fig. 5. Here, E_s is the peak amplitude of the signal potential, as indicated by equation 3, and $g_{\max}'$ and $i_{\max}'$ are the values used in equations 11 and 17 and pictured graphically in Fig. 2. This ratio is important in cases where the injection frequency and difference frequency are not widely separate, since the filter which usually follows the mixer must be designed to pass one and reject the other.

The curves of Fig. 5 exhibit sharp corners at $e_{im}=0$ and steep slopes as e_{im} approaches $-1/2$ because of the discontinuities in the i' and g' curves at these points. In curves drawn from measurements in actual circuits, it will appear subsequently that these features are missing, as should be expected. The theoretical and measured curves show reasonably close agreement, however, for values of e_{im} well removed from -0.5 and 0.

Injection Modulation Component

The injection source always has at least a slight modulation in its amplitude because of noise, power-frequency ripple, microphonics, etc. When the injection potential in a heterodyne mixer becomes very large, these variations may become the limiting factor in mixer sensitivity. The d-c component I_{dc} of the output current of the heterodyne mixer is a function of the amplitude of the injection potential. Consequently, when the injection potential varies in amplitude in accordance with noise, ripple, microphics, etc., the d-c output current varies in a corresponding manner. This variation in injection potential amplitude is equivalent to an equal variation in the peak value e_{im} while E_{dc} remains constant. Consequently, the relative amplitude of the injection modulation-frequency component in the output current can be expressed

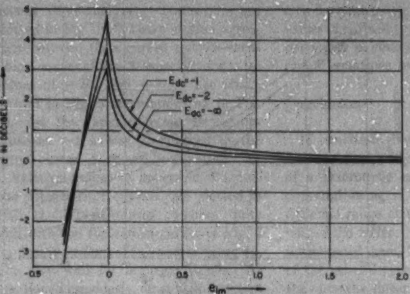


Fig. 7. Theoretical value of the figure of merit α, in decibels; this is the ratio of the relative conversion transconductance to the relative amplitude of the injection modulation component of the output current, as given by equations 27–30

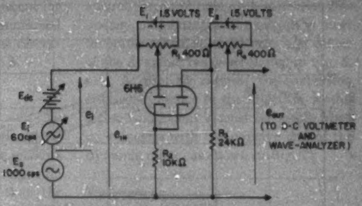


Fig. 8. An experimental circuit used to illustrate the application of the piecewise-linear method of analysis

in terms of the slope of a curve of the d-c current plotted against e_{im}, for various fixed values of E_{dc}.

In equation 15, the d-c component of i' is $I_{dc}=B_0/2$. The formula for this Fourier coefficient is

$$I_{dc}=B_0/2=(1/\pi)\int_0^\pi i'd(\omega_i t) \tag{22}$$

From inspection of Fig. 2, it is evident that i' is equal to a constant value $i'_{\max}$ between the limits $\omega_i t=0$ and $\omega_i t=\theta_3$, and equal to zero from $\omega_i t=\theta_3$ to $\omega_i t=\pi$, so that equation 22 reduces to

$$I_{dc}=B_0/2=(1/\pi)\int_0^{\theta_3} i'_{\max}d(\omega_i t) = (\theta_3/\pi)i'_{\max} \tag{23}$$

In general, E_{dc} is a large negative quantity. It will be assumed here that its largest permitted value in the positive direction is $-1/2$, in which case the injection potential swing is symmetrically placed with respect to the circuit characteristics. For this limiting case, $\theta_3=\pi/2$, so that the corresponding limiting value of $B_0/2$ is $(I_{dc})_{\max}=(1/2)i'_{\max}$. The relative value $(I_{dc})_{\text{rel}}$ of the d-c current component I_{dc} is therefore equal to

$$(I_{dc})_{\text{rel}}=\frac{I_{dc}}{(I_{dc})_{\max}}=\frac{I_{dc}}{(1/2)i_{\max}'}=\frac{\theta_3}{\pi/2} \tag{24}$$

A graph of this quantity is given in Fig. 6 as a function of e_{im}, with E_{dc} as a parameter.

It is evident in Fig. 6 that all curves approach zero very rapidly as e_{im} approaches $-1/2$. This is due to the discontinuity in i'. This would lead to conclusion that the injection modulation component, which is proportional to the slope of the curves, as mentioned above, should have a very large value as e_{im} approaches $-1/2$. However, in any practical circuit, a sharp discontinuity in i' such as that shown by the dashed line in Fig. 1 does not exist; the actual behavior is more like that represented by the solid line in the same figure. Consequently, steep slopes in the d-c curves are absent in practice, and the curves of Fig. 6 are not very realistic for negative values of e_{im}. Actually, this is not very important in practice, because high-injection heterodyne mixers invariably are driven vigorously enough so that e_{im} goes well into the positive region, and it will be shown subsequently that there is good agreement between experiment and the theoretical curves of Fig. 6 in the positive range of e_{im}.

From elementary trigonometry, we find that the angle θ_3 in equation 24 is

$$\theta_3=\operatorname{arcsec}\frac{-E_{dc}+e_{im}}{-E_{dc}-1/2} \tag{25}$$

The derivative of equation 25 with respect to e_{im} is

$$\frac{d}{de_{im}}\theta_3=\frac{(-E_{dc}-1/2)}{(-E_{dc}+e_{im})\times\sqrt{(-E_{dc}+e_{im})^2-(-E_{dc}-1/2)^2}} \tag{26}$$

which is a measure of the relative amplitude of the injection modulation-frequency component. The ratio of the relative conversion transconductance as given by equations 13 and 14 to the relative amplitude of the injection modulation-frequency component given by equation 26 is a figure of merit for the mixer which requires close attention when any injection modulation component falls within the passband of the difference-frequency filter following the mixer. The symbol α will be used here to represent figure of merit. When e_{im} lies between $-1/2$ and 0, α is found by dividing equation 14 by equation 26, with the following result:

$$\alpha=\frac{\sqrt{e_{im}^2-2E_{dc}e_{im}-2E_{dc}-1}\,\sqrt{e_{im}^2-2E_{dc}e_{im}-E_{dc}-1/4}}{(-E_{dc}-1/2)} \tag{27}$$

When e_{im} is positive, α is found by dividing equation 13 by equation 26, with the result:

$$\alpha=\frac{\left[\sqrt{e_{im}^2-2E_{dc}e_{im}-2E_{dc}-1}-\sqrt{e_{im}^2-2E_{dc}e_{im}}\right]\sqrt{e_{im}^2-2E_{dc}e_{im}-E_{dc}-1/4}}{(-E_{dc}-1/2)} \tag{28}$$

These values of α, expressed in decibels, are plotted in Fig. 7 as functions of e_{im}, with E_{dc} as a parameter. The values evidently change very little as func-

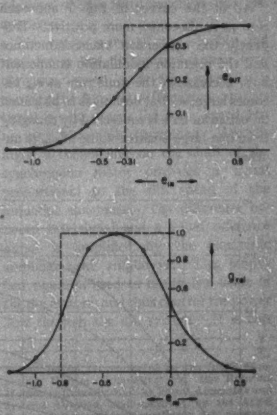


Fig. 9. Measured values of the instantaneous output potential and incremental relative transfer conductance for the circuit shown in Fig. 8, as functions of the instantaneous total injection potential

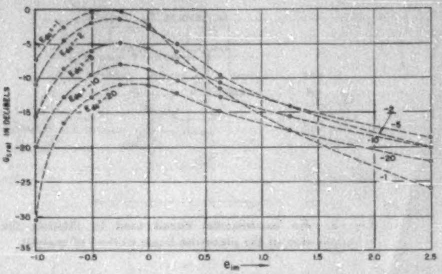


Fig. 10. Measured value of relative conversion transconductance for the circuit of Fig. 8; this may be compared with the theoretical values in Fig. 3

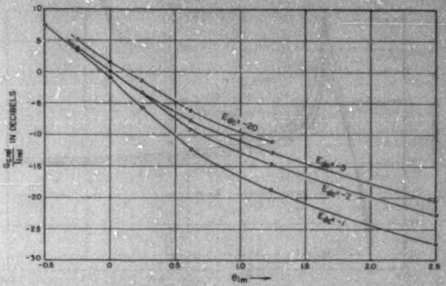


Fig. 12. The ratio of the measured values of G_{crel} and I_{1rel} for the circuit of Fig. 8, expressed in decibels; these curves may be compared with the theoretical curves of Fig. 5

tions of E_{dc}, so that only three curves, drawn for $E_{dc}=-1$, -2, and $-\infty$, are sufficient to describe them. In the limiting case where $E_{dc}=-\infty$, equations 27 and 28 reduce, respectively, to

$$\alpha=\sqrt{2(e_{im}+1)(2e_{im}+1)} \tag{29}$$

$$\alpha=\left(\sqrt{2(e_{im}+1)}-\sqrt{2e_{im}}\right)\sqrt{2e_{im}+1} \tag{30}$$

The sharp points in the curves at $e_{im}=0$ are due to the discontinuity in g' at this point, while the large attentuation as e_{im} approaches $-1/2$ is due to the discontinuity in i', as mentioned previously. In a practical circuit, the sharp points are rounded off and the attenuation is less severe.

All of the curves in Fig. 7 approach zero as e_{im} becomes more positive. Evidently the conversion transconductance and the injection modulation component both decrease at the same rate at e_{im} becomes large, so that nothing is to be gained or lost as far as α is concerned by changing from one large positive value of e_{im} to another.

Experimental Results

As an illustration the application of the piecewise-linear method of analysis, the experimental audio frequency circuit shown in Fig. 8 is used. This particular circuit is not used generally in heterodyne mixer applications and serves here merely as a laboratory analog for the class of mixers upon which this report is based. Audio frequencies are used to illustrate the method because the wave analyzer which has been made available for this investigation is of the audio-frequency type. The signal and injection frequencies are 1,000 and 60 cycles per second, respectively. The output current through the 24,000-ohm resistor, R_3, is proportional to the output potential, the components of which are measured by means of a high-resistance d-c voltmeter and the audio-frequency wave-analyzer just mentioned. The measured values of the d-c output voltage and the relative transconductance are plotted as solid lines in Fig. 9, as functions of the instantaneous total value of the input potential. The dashed lines in Fig. 9 are piecewise-linear approximations, determined in the manner described in connection with Fig. 1, to which Fig. 9 corresponds. For convenience in relating the measured values to the theoretical values, the 1.5-volt d-c source E_1 and the potentiometer R_1 are included in the input circuit and the potentiometer is adjusted to make the right-hand step in g' fall exactly zero on the e_t scale. The d-c potential appearing across the left-hand section of R_1 acts as a counter bias to the internal thermionic potential of the left-hand diode. The difference between the gating levels of the circuit depends upon the cathode temperature, which is held constant by holding the heater potential constant. The d-c source E_2 and the potentiometer R_4 in the ouptut circuit are used to adjust the d-c level of the output potential to the desired position.

By applying a graphical integration method to the g curve, the appropriate dynamic range of the g' curve is found to be 0.80 volt. This value then serves as the basis of normalization for the input voltage components E_{dc} and e_{im} where they appear in Figs. 10, 11, and 12, so that these graphs of the measured values can be compared directly with the theoretical curves in Figs. 3, 5, and 6. In Fig. 9, the abscissa represents the actual measured input potential, not the normalized value. When R_1 is adjusted so that right-hand step in g' falls exactly at zero, the left-hand step then falls at -0.80 volt (normalized value -1). Another graphical integration, this time applied to the i curve, then establishes the correct location for the step in i' at -0.31 volt (normalized value about -0.39).

Relative values of conversion transconductance for the circuit of Fig. 8 are measured directly with the audio-fre-

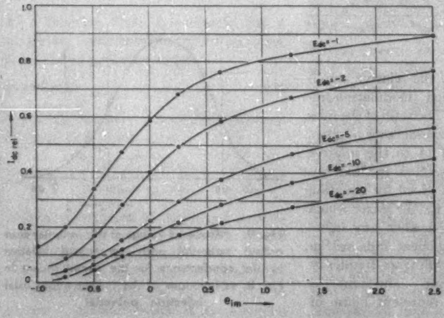


Fig. 11. Measured values of relative d-c current in the output of the circuit shown in Fig. 8; these curves may be compared with the theoretical curves in Fig. 6

quency wave-analyzer. The results are plotted in Fig. 10. There is good agreement between these curves and the theoretical curves of Fig. 3, especially for large values of $-E_{sb}$ and $+e_{im}$. The sharp corners in Fig. 3 are rounded off in Fig. 10, and the rapid decrease as e_{im} approaches −1 is less pronounced in the actual circuit, as should be expected.

Relative values of measured d-c output potential (or current) are plotted in Fig. 11. These curves compare quite well with the theoretical curves of Fig. 6 for positive values of e_{im}, but, as expected, not very well for negative values. As mentioned previously, most high-injection mixers operate in the positive range of e_{im}, where the theoretical curves of Fig. 6 agree well with experiment.

Fig. 12 shows the ratio of the measured values of $G_{c\,rel}$ and $I_{i\,rel}$ for the circuit of Fig. 8, expressed in decibels. These curves compare favorably with the theoretical values plotted in Fig. 5, except for values of e_{im} near 0 or −0.5, as mentioned previously in connection with the discussion of equations 19–21 and Fig. 5.

Conclusions

It has been shown that a piecewise-linear method is adequate for the analysis of a heterodyne mixer if the injection potential swing is much large than the dynamic range of the injection terminal pair of the mixer. In the class of mixers investigated here, having characteristics of the general type illustrated in Fig. 1, the piecewise-linear method has been used to demonstrate three things: 1. that the largest conversion transconductance is obtained when the maximum instantaneous total injection potential (e_{im}) coincides with the positive end of the dynamic range of the mixer, as shown in Fig. 3; 2. that the ratio of the conversion transconductance to the amplitude of the injection frequency component in the output is most favorable when e_{im} falls near the center of the dynamic range, as shown in Fig. 5, and, 3. that the ratio of the conversion transconductance to the injection modulation component is most favorable when e_{im} coincides with the positive end of the dynamic range, as shown in Fig. 7.

References

1. The Operation of Frequency Converters and Mixers for Superheterodyne Reception, E. W. Herold. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 30, Feb. 1942, p. 84.

2. Electron-Tube Circuits (book), S. Seely. McGraw-Hill Book Company, Inc., New York, N. Y., 1950, pp. 359–61.

3. Microwave Receivers (book), S. N. Van Voorhis. *Ibid.*, 1948, pp. 138–48.

4. Copper Oxide Modulators in Carrier Systems, Caruthers. *Bell System Technical Journal*, New York, N. Y., vol. 18, April 1939, pp. 315–37.

5. Equivalent Modulator Circuits, Peterson, Hussey. *Ibid.*, vol. 18, Jan. 1939, pp. 32–37.

Analog Computer Study of Wind-Tunnel Drive

K. G. BLACK
ASSOCIATE MEMBER AIEE

R. J. NOORDA
ASSOCIATE MEMBER AIEE

PRIOR to the installation of a large wind-tunnel drive rated 110,000 hp (horsepower) continuously and 132,000 hp for one hour, an analog computer study was made of the drive and its control. Although considerable previous experience had been acquired in the design of this type of drive, two factors made it desirable to examine the system on the computer. One factor is the large power usage of the drive and the length of time involved in acceleration and deceleration at a rate acceptable to the electrical utility. These considerations prohibited excessive tune-up adjustments and testing. The other factor is the transient current in the d-c machine during switching. It was necessary to determine that this current would not exceed circuit-breaker settings.

Description of Drive

The wind-tunnel drive studied is an adjustable speed drive in which continuous speed control is obtained from zero speed to 485 rpm, or 94.4% of synchronous speed. It consists of three 14-pole wound-rotor induction machines coupled in tandem on a single shaft. The smaller machine is a doubly fed machine which is part of a Clymer-drive system, while the two larger machines are controlled by variable rotor resistors. The drive and control system are shown in simplified form in Fig. 1. Fig. 2 shows the wind tunnel in which this drive is used. A more detailed description and the advantages of this type of drive are given in previously published papers.[1–3]

All three induction machines have liquid rheostats in the rotor circuits. In parallel with the rheostat in the rotor of the 27,000-hp machine is connected the stator of a synchronous motor. The rheostats for the 41,500-hp machines control the division of load between the three induction machines while the rheostat for the 27,000-hp machine shares load with the synchronous motor. The synchronous motor drives two d-c generators which in turn feed two d-c motors. A synchronous generator coupled to the d-c motors feeds power back into the a-c system. The rheostats are fairly slow acting and may be considered to maintain steady-state load division while the high-accuracy speed control is provided by the d-c system and its regulator.

At fan speeds up to 35%, the 27,000-hp machine drives alone. At this speed one of the 41,500-hp machines is connected to the line, and at 46.5% speed the second 41,500-hp machine is added. The liquid rheostats of both 41,500-hp machines are maintained at the maximum resistance value until the transient from the addition of the third machine has disappeared.

The rotor speed of the doubly fed machine is the difference between synchronous speed and slip speed. By exciting the rotor at a frequency proportional to slip speed, the rotor speed can be very accurately controlled. Accordingly, a tachometer on the shaft of the synchronous motor provides a signal proportional to the slip speed of the induction machines. The difference between this signal and a speed reference signal is amplified by the electronic amplifier and amplidyne. The amplified signal is used to excite the fields of the d-c motors, and thus control the power flow through the d-c system. It is this

---

Paper 57-1006, recommended by the AIEE Computing Devices Committee and the Rotating Machinery Committee of the Power Apparatus and Systems Division, and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted June 20, 1957; made available for printing August 12, 1957.

K. G. Black and R. J. Noorda are with the General Electric Company, Schenectady, N. Y.

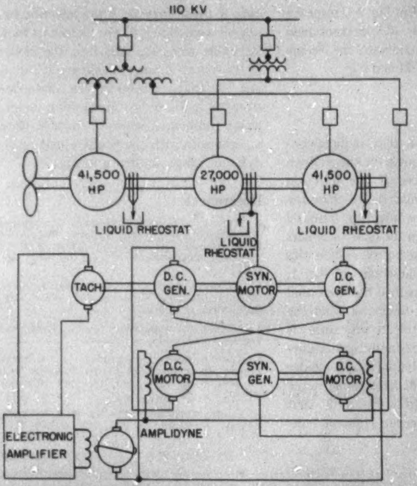

110 KV
41,500 HP
27,000 HP
41,500 HP
LIQUID RHEOSTAT
LIQUID RHEOSTAT
LIQUID RHEOSTAT
TACH
D.C. GEN.
SYN. MOTOR
D.C. GEN.
D.C. MOTOR
SYN. GEN.
D.C. MOTOR
ELECTRONIC AMPLIFIER
AMPLIDYNE

Fig. 1. Drive and control

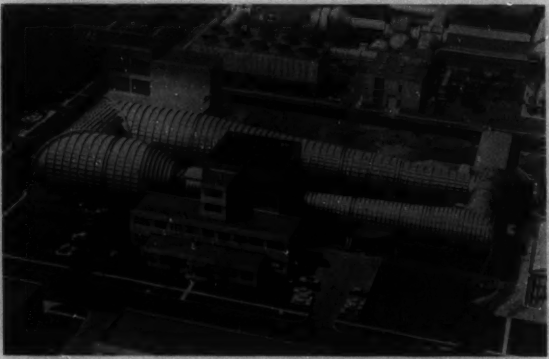


Fig. 2 (below). The National Advisory Committee for Aeronautics 14-foot transonic wind tunnel at Moffet Field, California

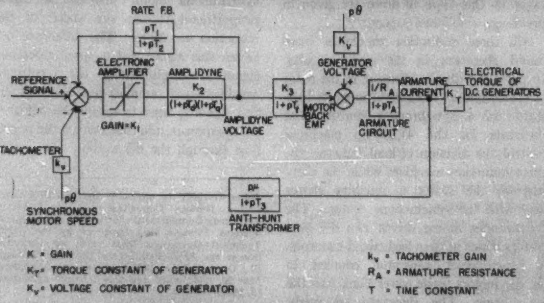

RATE F.B.
$\frac{pT_1}{1+pT_2}$
ELECTRONIC AMPLIFIER
AMPLIDYNE
REFERENCE SIGNAL
GAIN = K_1
$\frac{K_2}{(1+pT_d)(1+pT_q)}$
AMPLIDYNE VOLTAGE
$\frac{K_3}{1+pT_f}$
MOTOR BACK EMF
$p\theta$
K_v
GENERATOR VOLTAGE
$\frac{1/R_A}{1+pT_A}$
ARMATURE CIRCUIT
ARMATURE CURRENT
K_T
ELECTRICAL TORQUE OF D.C. GENERATORS
TACHOMETER
k_v
$p\theta$
SYNCHRONOUS MOTOR SPEED
$\frac{p\mu}{1+pT_3}$
ANTI-HUNT TRANSFORMER
K = GAIN
K_T = TORQUE CONSTANT OF GENERATOR
K_v = VOLTAGE CONSTANT OF GENERATOR
k_v = TACHOMETER GAIN
R_A = ARMATURE RESISTANCE
T = TIME CONSTANT

Fig. 3. Block diagram of d-c control system

power taken from the rotor of the doubly fed machine which effects the control of fan speed.

System Representation

The various components of the drive and control were analyzed and an analog computer representation developed. The doubly fed machine was represented by use of the equations derived in Appendix I. The synchronous machine was represented by the well-known equations of Appendix II, and the d-c system by the block diagram of Fig. 3. The schematic of the computer circuit for representation of the doubly fed machine and synchronous motor is given in Fig. 4. The load torque on the drive caused by air resistance was known to be approximately proportional to the square of the fan speed, and it was so represented. Examination of the parameters of the 41,500-hp induction machines revealed that the electrical time constants were small compared with the transients in the mechanical system. These induction machines were accordingly represented by their steady-state speed-torque curves.

The liquid rheostat regulators were omitted from the study. It was considered that they were slower acting than the d-c loop and could be neglected. Any error introduced by this omission would be on the pessimistic side, and thus permissible. This did, of course, introduce an element of the small oscillation type of study which otherwise had been avoided. As a result, it was necessary to calculate by hand the rotor circuit resistances for each fan speed, and then examine operation of the system around that speed. This was not true, of course, for the 41,500-hp machines at speeds below 46.5% of synchronous speed, because then their rheostat regulators were inoperative.

The synchronous generator was not included in the computer representation. It was considered to be connected to an infinite bus and, thus, would not affect the drive and control.

Results of Study

The representation of the system as developed in this paper was placed on the analog computer. Then, the stability of the complete drive was examined throughout the operating speed range. Such modifications as needed were made in the regulator circuit to provide a very stable system. While operating the drive at a speed of 36% of synchronous, the first 41,500-hp machine was added

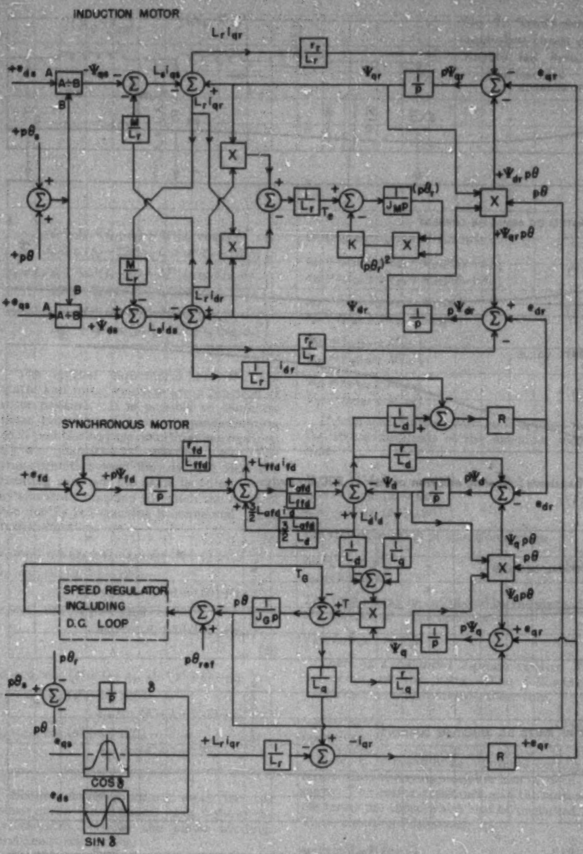


Fig. 4. Computer schematic

and the effect on the system observed. Fig. 5 shows the transients resulting therefrom. Note that in spite of the omission in the computer representation of the rheostat regulator for the 27,000-hp machine, the direct-current did not exceed the breaker setting. In actual operation, with the rheostat regulator beginning to act before the maximum direct current was reached, the peak value would be less. Fig. 6 shows the transients resulting from the addition of the second 41,500-hp machine at a speed of 47% of synchronous. As would be expected, the transients are less severe than those of Fig. 5.

As customary with a complex computer representation, precautions were taken to preclude the possibility of errors in the computer setup. One method was to compare the steady-state operating condition as obtained from the computer with values obtained by hand calculation. Appendix IV presents the steady-state equations of the doubly fed machine from which the equivalent circuit of Fig. 7 is obtained.

The appendix also outlines the method of calculation. The drive and controls were installed in the wind tunnel and were operated as indicated by the computer study.

Appendix 1. Derivation of Equations of Doubly Fed Induction Motor

Nomenclature

DOUBLY FED MACHINE

e_a, e_b, e_c = phase terminal voltages of stator
e_{dr}, e_{qr}, e_{or} = rotor voltages in terms of components along *d*, *q*, *o* axes
e_{ds}, e_{qs}, e_{os} = stator voltages in terms of components along *d*, *q*, *o* axes
e_1, e_2, e_3 = phase terminal voltages of rotor
i_a, i_b, i_c = phase currents in stator; current flowing into stator is taken as positive
i_{dR}, i_{qR} = current through rotor resistor in terms of components along the *d*, *q* axes
i_{dr}, i_{qr} = currents of rotor windings in terms of components along the *d*, *q* axes
i_{ds}, i_{qs} = currents of stator windings in terms of components along the *d*, *q* axes
i_1, i_2, i_3 = phase currents in rotor; current flowing into rotor is taken as positive
$L_r = L_r' - M_r'$ = rotor self-inductance
L_r' = self-inductance of each rotor winding
$L_s = L_s' - M_s'$ = stator self-inductance
L_s' = self-inductance of each stator winding
$M = 3/2\ M'$ = mutual inductance between rotor and stator
M' = coefficient of mutual inductance between rotor and stator windings
M_r' = mutual inductance between rotor windings
M_s' = mutual inductance between stator windings
$p = d/dt$ = derivative operator
R = resistance of rotor resistor
r_r = resistance of each rotor winding
r_s = resistance of each stator winding
T_e = torque on shaft from machine
V = rms value of 3-phase voltage applied to stator
$\delta = \theta - \theta'$ = angle between *d* axis and *d'* axis
Ψ_a, Ψ_b, Ψ_c = flux linkages of stator windings
Ψ_{dr}, Ψ_{qr} = flux linkages of rotor windings in terms of components along the *d*, *q* axes
Ψ_{ds}, Ψ_{qs} = flux linkages of stator windings in terms of components along the *d*, *q* axes
Ψ_1, Ψ_2, Ψ_3 = flux linkages of rotor windings
$\theta' = \theta_s - \theta_r$ = slip angle (angle between the rotor and an axis rotating at slip speed with respect to the rotor)
θ_r = angle between phase 1 winding of rotor and *a*-phase winding of stator
θ_s = angle between axis of phase *a* and axis rotating at synchronous speed

Equations for the performance of the doubly fed machine have previously been developed and may be found in the literature.[4-6] Although equations 15 of reference 5 are well suited for differential analyzer solution, for the present case, in which a synchronous machine is connected to the rotor of the induction machine, the equations of this appendix provide additional simplification. Their derivation follows the well-known method of Park.[7]

The following assumptions are made regarding the machine:

1. The rotor and stator windings are balanced.

2. The mutual inductance between any stator winding and any rotor winding is a

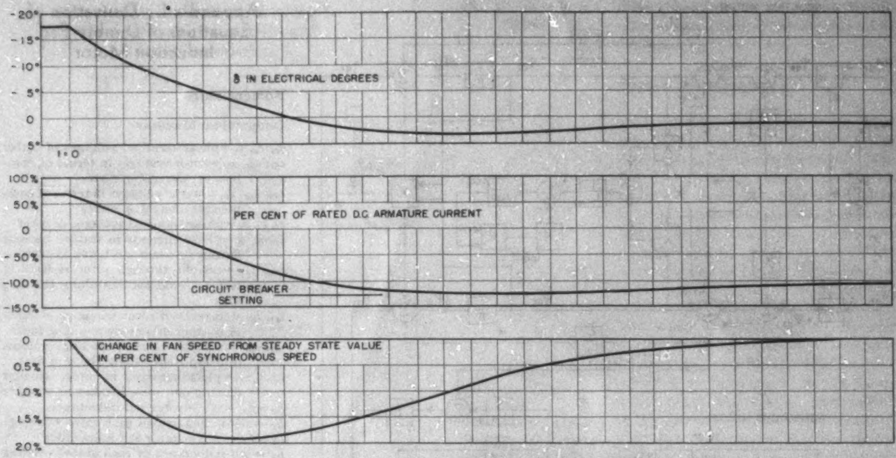


Fig. 5. Transients following addition of first 41,500-hp motor

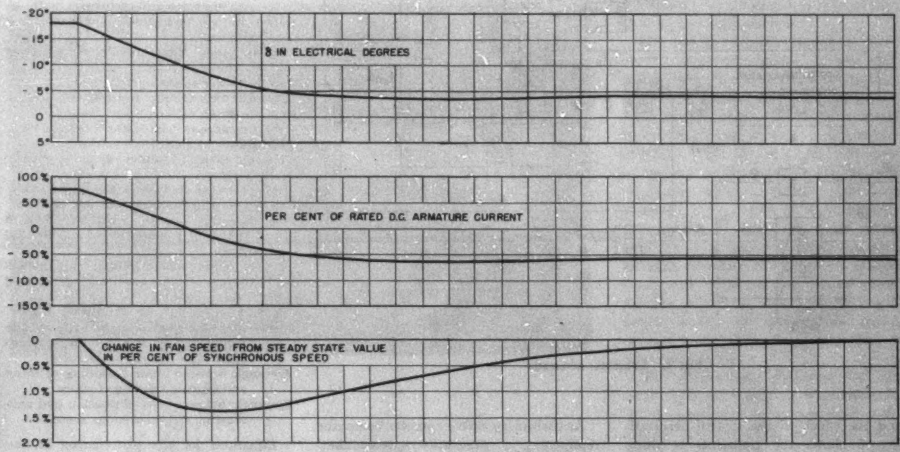


Fig. 6. Transients following addition of second 41,500-hp motor

cosinusoidal function of the electrical angle between the axes of the two windings.

3. The rotor and stator are smooth and the self-inductances of all windings are independent of rotor position.

4. The effects of saturation, hysteresis, and eddy currents are neglected.

Let the several windings of the machine be replaced by three windings on the rotor and three windings on the stator. Fig. 8 shows the relative position of the windings. The rotor windings rotate at a speed of $p\theta_r$. The following equations may be written:

For the stator windings

$$\begin{aligned} e_a &= r_s i_a + p\Psi_a \\ e_b &= r_s i_b + p\Psi_b \\ e_c &= r_s i_c + p\Psi_c \end{aligned} \tag{1}$$

and for the rotor windings

$$\begin{aligned} e_1 &= r_r i_1 + p\Psi_1 \\ e_2 &= r_r i_2 + p\Psi_2 \\ e_3 &= r_r i_3 + p\Psi_3 \end{aligned} \tag{2}$$

As a result of the assumptions about the inductances of the machine, the fluxes become

$$\left.\begin{aligned} \Psi_a &= L_s' i_a + M_s'(i_b + i_c) + M'[i_1 \cos\theta_r + i_2 \cos(\theta_r + 120) + i_3 \cos(\theta_r - 120)] \\ \Psi_b &= L_s' i_b + M_s'(i_c + i_a) + M'[i_1 \cos(\theta_r - 120) + i_2 \cos\theta_r + i_3 \cos(\theta_r + 120)] \\ \Psi_c &= L_s' i_c + M_s'(i_a + i_b) + M'[i_1 \cos(\theta_r + 120) + i_2 \cos(\theta_r - 120) + i_3 \cos\theta_r] \end{aligned}\right\} \tag{3}$$

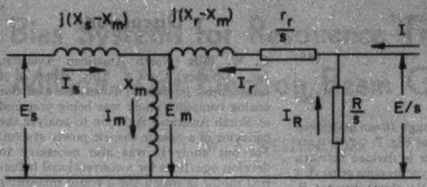


Fig. 7. Steady-state equivalent circuit of doubly fed induction machine

$$\left.\begin{aligned}\Psi_1&=L_r'i_1+M_r'(i_2+i_3)+M'[i_a\cos\theta_r+\\&\qquad i_b\cos(\theta_r-120)+i_c\cos(\theta_r+120)]\\\Psi_2&=L_r'i_2+M_r'(i_3+i_1)+M'[i_a\cos(\theta_r+\\&\qquad 120)+i_b\cos\theta_r+i_c\cos(\theta_r-120)]\\\Psi_3&=L_r'i_3+M_r'(i_1+i_2)+M'[i_a\cos(\theta_r-\\&\qquad 120)+i_b\cos(\theta_r+120)+i_c\cos\theta_r]\end{aligned}\right\} \tag{4}$$

The mutual inductances between the stator and rotor windings are a function of rotor position. It is possible to eliminate these functions of θ_r by a transformation such that the 3-phase quantities are replaced by d-c quantities on rotating axes. The transformation used here is such that the rotating axes are the same as those of the synchronous machine to which the rotor of the doubly fed machine is connected. The transformations are

$$\left.\begin{aligned}e_a&=e_{ds}\cos(\theta_s+\delta)-e_{qs}\sin(\theta_s+\delta)+e_{0s}\\e_b&=e_{ds}\cos(\theta_s+\delta-120)-\\&\qquad e_{qs}\sin(\theta_s+\delta-120)+e_{0s}\\e_c&=e_{ds}\cos(\theta_s+\delta+120)-\\&\qquad e_{qs}\sin(\theta_s+\delta+120)+e_{0s}\end{aligned}\right\} \tag{5}$$

$$\left.\begin{aligned}e_1&=e_{dr}\cos(\theta'+\delta)-e_{qr}\sin(\theta'+\delta)+e_{0r}\\e_2&=e_{dr}\cos(\theta'+\delta-120)-\\&\qquad e_{qr}\sin(\theta'+\delta-120)+e_{0r}\\e_3&=e_{dr}\cos(\theta'+\delta+120)-\\&\qquad e_{qr}\sin(\theta'+\delta+120)+e_{0r}\end{aligned}\right\} \tag{6}$$

Similar transformations exist for the phase currents and fluxes. Fig. 9 shows the relationship between the phase windings and the rotating axes.

In this study, only balanced 3-phase operation is of interest. Hence, the zero-phase-sequence components may be omitted. If the transformation equations are substituted into equations 1 through 4, the resulting equations of the machine are

$$\left.\begin{aligned}e_{ds}&=r_si_{ds}+p\Psi_{ds}-\Psi_{qs}p(\theta_s+\delta)\\e_{qs}&=r_si_{qs}+p\Psi_{qs}+\Psi_{ds}p(\theta_s+\delta)\end{aligned}\right\} \tag{7}$$

$$\left.\begin{aligned}e_{dr}&=r_ri_{dr}+p\Psi_{dr}-\Psi_{qr}p\theta\\e_{qr}&=r_ri_{qr}+p\Psi_{qr}+\Psi_{dr}p\theta\end{aligned}\right\} \tag{8}$$

$$\left.\begin{aligned}\Psi_{ds}&=L_si_{ds}+Mi_{dr}\\\Psi_{qs}&=L_si_{qs}+Mi_{qr}\end{aligned}\right\} \tag{9}$$

$$\left.\begin{aligned}\Psi_{dr}&=L_ri_{dr}+Mi_{ds}\\\Psi_{qr}&=L_ri_{qr}+Mi_{qs}\end{aligned}\right\} \tag{10}$$

If balanced 3-phase voltage is applied to the stator terminals of the machine, such that

$$\left.\begin{aligned}e_a&=-V\sin\theta_s\\e_b&=-V\sin(\theta_s-120)\\e_c&=-V\sin(\theta_s+120)\end{aligned}\right\} \tag{11}$$

the transformation of equation 5 gives

$$\left.\begin{aligned}e_{ds}&=V\sin\delta\\e_{qs}&=V\cos\delta\end{aligned}\right\} \tag{12}$$

The torque may be calculated by the method of reference 8. The result is

$$T_e=i_{dr}\Psi_{qr}-i_{qr}\Psi_{dr} \tag{13}$$

There is a balanced 3-phase resistor R connected to the rotor. The following additional equations result therefrom.

$$\begin{aligned}i_{dR}&=-e_{dr}/R\\i_{qR}&=-e_{qr}/R\end{aligned} \tag{14}$$

Now, it is possible to eliminate certain terms without any appreciable loss of accuracy. The stator resistance and the stator $p\Psi$ terms are those which can be neglected. Thus, equation 7 becomes

$$\begin{aligned}e_{ds}&=-\Psi_{qs}p(\theta_s+\delta)\\e_{qs}&=\Psi_{ds}p(\theta_s+\delta)\end{aligned} \tag{15}$$

The machine is now completely represented by equations 8 through 10 and 12 through 15.

Appendix II. Equations of Synchronous Machine and Interconnection Equations

Nomenclature

SYNCHRONOUS MOTOR

e_d = direct-axis-armature terminal voltage
e_{fd} = field voltage
e_q = quadrature-axis-armature terminal voltage
i_d = direct-axis-armature current
i_{fd} = field current
i_q = quadrature-axis-armature current
L_{afd} = mutual-inductance coefficient between armature flux linkage and rotor current
L_d = direct-axis-armature inductance
L_{ffd} = field inductance
L_q = quadrature-axis-armature inductance
r = armature resistance
r_f = field resistance
T = electrical torque
Ψ_d = direct-axis-armature flux linkages
Ψ_{fd} = field flux linkages
Ψ_q = quadrature-axis-armature flux linkages

The equations used to represent the synchronous machine are those of Park.[7] Their derivation is given in the literature[7,8] and is not repeated here. A simplification of these equations is effected by the omission of the amortisseur windings. This will provide conservative results, since the resulting computer representation will exhibit less damping than actually exists. Subject to this simplification and if only balanced 3-phase operation is encountered, Park's equations are

$$e_d=p\Psi_d-\Psi_qp\theta-ri_d \tag{16}$$

$$e_q=p\Psi_q+\Psi_dp\theta-ri_q \tag{17}$$

$$e_{fd}=r_fi_{fd}+p\Psi_{fd} \tag{18}$$

$$\Psi_d=-L_di_d+L_{afd}i_{fd} \tag{19}$$

$$\Psi_q=-L_qi_q \tag{20}$$

$$\Psi_{fd}=-3/2L_{afd}i_d+L_{ffd}i_{fd} \tag{21}$$

$$T=i_q\Psi_d-i_d\Psi_q \tag{22}$$

The interconnection between the doubly fed induction motor and the synchronous machine requires the following equations.

$$e_d=e_{dr} \tag{23}$$

$$e_q=e_{qr} \tag{24}$$

$$i_{dr}=i_d+i_{dR} \tag{25}$$

$$i_{qr}=i_q+i_{qR} \tag{26}$$

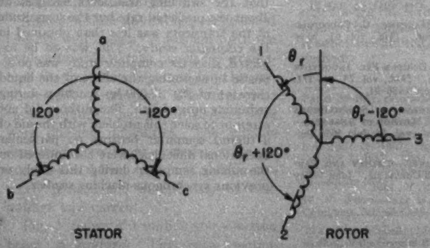


Fig. 8. Relation of rotor and stator windings of doubly fed machine

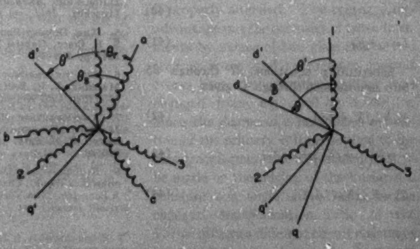


Fig. 9. Relation of rotating axes

Appendix III. Block Diagram of Control System

The control system consists of the tachometer generator, electronic amplifier, amplidyne, d-c motors, and d-c generators, as shown in Fig. 1. Since the equations of these components are well known, no derivation of them is given here.

The electronic amplifier contains no phase-shifting networks, so it can be represented by a constant gain and saturation limits. Such feedback as exists around the amplidyne has only a small effect, so the amplidyne can be represented by a gain and two time constants. Combination of the components results in the block diagram of Fig. 3. For stabilization, rate feedback from amplidyne voltage and armature current is used.

Appendix IV. Steady-State Operation

In steady-state operation, the flux derivative is zero as well as $p\delta$. In addition, $p\theta_s=\omega$, $p\theta=s\omega$, where $s=$slip, and $\omega L=X$. Thus, in the steady state, equations 8 through 10 and equations 12 through 15 become

$$e_{ds}=-X_s i_{qs}-X_m i_{qr} \tag{27}$$

$$e_{qs}=X_s i_{ds}+X_m i_{dr} \tag{28}$$

$$e_d=r_r i_{dr}-(X_r i_{qr}+X_m i_{qs})s \tag{29}$$

$$e_q=r_r i_{qr}+(X_r i_{dr}+X_m i_{ds})s \tag{30}$$

$$T_e=X_m(i_{dr}i_{qs}-i_{qr}i_{ds}) \tag{31}$$

$$e_d=-Ri_{dR} \tag{32}$$

$$e_q=-Ri_{qR} \tag{33}$$

$$i_{dr}=i_d+i_{dR} \tag{34}$$

$$i_{qr}=i_q+i_{qR} \tag{35}$$

Let

$$E_s=e_{ds}+je_{qs} \tag{36}$$

$$I_s=i_{ds}+ji_{qs} \tag{37}$$

$$I_R=i_{dR}+ji_{qR} \tag{38}$$

$$E=e_d+je_q \tag{39}$$

$$I_r=i_{dr}+ji_{qr} \tag{40}$$

$$I=i_d+ji_q \tag{41}$$

Substituting equations 27 through 35 into equations 36 through 41 gives

$$E_s=I_sjX_s+I_rjX_m \tag{42}$$

$$\frac{E}{s}=I_r\left(\frac{r_r}{s}+jX_r\right)+I_sjX_m \tag{43}$$

$$T_e=R_e[jX_m(I_rI_s^*)] \tag{44}$$

where R_e means real part of

$$\frac{E}{s}=-\frac{R}{s}\ I_R \tag{45}$$

$$I_r=I+I_R \tag{46}$$

From equations 42 through 46 can be derived the equivalent circuit of Fig. 7.

The total slip power is divided between that delivered to the synchronous generator and that dissipated in the shunt rheostat R. For a given power P_{SG}, delivered by the slip generator, the corresponding value of R can be calculated.

$$P_{SG}=R_e(EI^*) \tag{47}$$

But

$$I=I_r-I_R=i_{dr}+ji_{qr}+\frac{e_d}{R}+j\frac{e_q}{R}$$

and

$$I^*=\left(i_{dr}+\frac{e_d}{R}\right)-j\left(i_{qr}+\frac{e_q}{R}\right)$$

Thus

$$P_{SG}=\left(i_{dr}+\frac{e_d}{R}\right)e_d+\left(i_{qr}+\frac{e_q}{R}\right)e_q \tag{48}$$

or

$$R=\frac{|E|^2}{P_{SG}-(i_{dr}e_d+i_{qr}e_q)} \tag{49}$$

Finally

$$I=I_r-I_R$$

It is now possible to calculate the field excitation of the synchronous machine from the equations of Appendix II.

References

1. Large Adjustable-Speed Wind-Tunnel Drive, C. C. Clymer. *AIEE Transactions*, vol. 61, Mar. 1942, pp. 156–58, 456.

2. Variable-Speed Drive for United States Army Air Corps Wind Tunnel at Wright Field, Dayton, Ohio, A. D. Dickey, C. M. Laffoon, L. A. Kilgore. *Ibid.*, pp. 126–30.

3. A Study of the Modified Kramer or Asynchronous-Synchronous Cascade Variable-Speed Drive, M. M. Liwschitz, L. A. Kilgore. *Ibid.*, May 1942, pp. 255–60.

4. Damping and Synchronizing Torque of the Double-Fed Asynchronous Machine, M. M. Liwschitz. *Ibid.*, vol. 60, Oct. 1941, pp. 923–24.

5. The Doubly-Fed Machine, C. Concordia, S. B. Crary, Gabriel Kron. *Ibid.*, vol. 61, May 1942, pp. 286–89.

6. The Synchronous Double-Fed Induction Machine, R. E. Bedford. *Ibid.*, vol. 75, pt. III, 1956 (Feb. 1957 section), pp. 1486–91.

7. Two-Reaction Theory of Synchronous Machines, Generalized Method of Analysis—Part I, R. H. Park. *Ibid.*, vol. 48, July 1929, pp. 716–30.

8. Synchronous Machines—Theory and Performance (book), Charles Concordia. John Wiley & Sons, Inc., New York, N. Y., 1951.

Discussion

G. W. Bills (North American Aviation, Inc., Downey, Calif.): This paper is of considerable interest to us because a similar analog computer study was being prepared at North American Aviation to analyze the behavior of a missile electric power system. For our study it was also necessary to develop equations for a conventional induction motor in which the 3-phase quantities were replaced by d-c quantities on rotating axes. As pointed out by others, the use of synchronous- and induction-machine equations for transient conditions rather than simplified equivalent circuits is more accurate and desirable if the solutions of such equations can be obtained with the aid of digital or analog computers.

Approximately how many d-c amplifiers did you employ to represent your wind-tunnel system on the analog computer? What is the time scale in Figs. 5 and 6? Have your results been verified by operation of the wind tunnel?

Have you considered solving such problems digitally with the aid of an International Business Machines *704* or equivalent general-purpose digital computer? At North American Aviation we have solved synchronous-machine equations both by digital and analog methods, and from this experience we prefer digital methods, mainly because of operational difficulties with analog representations when many components are involved. Unfortunately, we have been unable to go ahead with the extension of the digital techniques to include induction motors as well as synchronous machines. However, we do not see any obstacles which prevent digital solution of such systems.

This commendable paper is evidence that many problems involving electric machinery can now be effectively analyzed with the help of existing computers. Heretofore, it has been necessary to so simplify such problems, in order that they could be handled by long-hand techniques or on mechanical differential analyzers, that the solutions were not too meaningful.

K. G. Black and **R. J. Noorda:** The authors appreciate the comments of Mr. Bills. The analog computer representation of the wind-tunnel system employed 64 amplifiers in addition to multipliers and function generators. The authors did not have the opportunity to compare actual wind-tunnel operation with the results of the computer study. However, it was reported to them that the switching transients decayed at about the predicted rate but the magnitude of the transients was less than obtained in the computer study. This was to be expected, since the computer study was pessimistic in neglecting the effect of the liquid rheostat of the 27,000-hp machine during switching operations. The authors did not attempt to solve this problem with the aid of a digital computer because no particular operational difficulties were encountered on the analog computer during this study or previous synchronous-machine studies.

Bias Systems for Resonance Transformer Million-Volt Electron Beam Generators

W. F. WESTENDORP
FELLOW AIEE

ELECTRON beam generating or X-ray equipment with the self-rectifying tube connected directly across the high-voltage resonance transformer[1–4] is a very reliable and economical type of high-voltage apparatus. The principal disadvantage of the spread of the electron energy during the useful half-cycle may be greatly reduced by the application of a periodically varying negative bias applied between a control electrode of the electron gun and the filament, so that current flow can take place only during the central portion of the half-wave of voltage.

The advantages of this type of bias with electron beam irradiation equipment are: 1. more nearly uniform penetration of all the electrons in the materials to be irradiated, with a consequent saving in power for a given minimum dose and a reduced overdose; 2. a more convenient sweep system for the electron beam across the conveyor belt since the deflection current in the sweep yoke can be nearly proportional to the desired deflection; and 3. the elimination of the low-energy electrons from the beam, reducing the heating produced in the metal foil window of the vacuum tube through which the beam emerges, which greatly facilitates the window-cooling problem.

Many advantages are also obtained by the use of this type of bias with X-ray tubes, such as increased X-ray output per milliampere of tube current; increased penetration of the radiation; better magnetic focusing of the nearly monoenergetic electron beam to a sharper and, if desired, smaller focal spot; and improved operating stability of the high-voltage multisection tube.

Although these advantages of the use of bias are well known,[6] it was not until the application of electron beams for irradiation that a serious demand for this improvement developed. The first bias circuits required high potentials, and there were phasing difficulties. The circuits described here avoid excessive voltages, and the final bias system is always correctly phased and independent of tube voltage and current.

Fig. 1 shows a 1-million-volt electron beam generator.[3] The cross section of a million-volt outfit is presented in Fig. 2. The coil stack of the resonance transformer surrounds the vacuum tube and is placed in the center of the steel tank which contains the gas insulation.[7] Originally Freon *12* was used, but at present the even better insulating gas sulphurhexafluoride is used at a pressure of 60 psi (pounds per square inch).[7] The coil is terminated in a rounded brass cap with radial slots to prevent excessive eddy currents. This cap serves as a housing for filament and bias circuits. The secondary coil resonates with its distributed and terminal capacity at 180 cps (cycles per second). The power is supplied by means of a relatively small primary on the bottom of the coil stack. The magnetic flux is returned along the inside of the tank wall through overlapping laminated silicon steel strips to a laminated silicon steel base under the coil stack. The tank diameter is 3 feet, the height $4^1/_2$ feet. Two- and 4-million-volt electron beam generators are of similar design with tank diameters respectively of $4^1/_2$ feet and 8 feet, and tank heights of 8 feet and 15 feet. A considerable literature has built up on the use of high-energy electrons in the fields of sterilization and chemistry.[8–13]

Fig. 3 shows from top to bottom the following:

Fig. 3(A): Over-all transformer or tube voltage which is a 180-cycle sine wave.

Fig. 3(B): The a-c charging current circulating in the transformer secondary which leads the voltage by 90 degrees and is of the order of 60, 130, and 350 ma (milliamperes) rms respectively for the 1-, 2- and 4,000,000-volt machines.

Fig. 3(C): The vacuum-tube electron current leaving the cathode without the use of bias. The sloping sides of the wave are space-charge limited; the flat top is emission limited.

Fig. 3(D): The desired waveform with bias. A conduction angle between 80 and 60 degrees is satisfactory for most cases. The average value of the vacuum-tube electron current may be from 1 to 6 ma.

For a conduction angle A radians the minimum voltage at start and finish of current flow is

$$V_{min} = V_{max} \cos 1/2A$$

The average voltage during conduction may be found by integration and is

$$V_{avg} = 2V_{max} \sin 1/2A/A$$

The crest value of the electron emission required from the filament is slightly more than

$$i_m = i_{avg} 2\pi/A$$

since the waveform is never perfectly rectangular. The average current i_{avg} is the conventional value of the electron current as read on a d-c instrument. Table I shows the effect of bias upon these three quantities.

Bias From Saturable Core Transformer

Fig. 4 shows the circuit diagram of a bias circuit employing a special "peaking" transformer. End turns, placed on top of the resonance transformer coil stack, are used to energize a 150-watt Variac. The output of the Variac passes through a phase adjustment resistor to the primary winding of a special transformer with two magnetic circuits *a* and *b*. Magnetic circuit *a* has a gap and does not saturate; *b* is closed and saturates twice per cycle. Primary winding *1* energizes both cores, winding *2* produces the well-known peaked voltage wave, and winding *3*, linking *a* and *b*, has the same sinusoidal waveform as winding *1*. The differences of the voltages of *2* and *3* constitute the desired bias voltage which is applied to the tube between focusing cup and filament.

This system was found to perform well if properly adjusted. The principal disadvantage was the necessity of two adjustments by means of two glass shafts, two stuffing boxes, and two remote controls every time the operating voltage was changed from, e.g., 500 to 1,000 kv. Since the degree of saturation is a function of the primary voltage applied to the bias transformer, one adjustment has to be made on the setting of the Variac if a constant bias angle is required. As the changed transformation ratio of the Variac reflects a different circuit resistance into the primary, a second adjustment is necessary in the phase correction re-

---

Paper 57-1057, recommended by the AIEE Nucleonics Committee and approved by the AIEE Technical Operations Department for publication without prior presentation. Manuscript submitted April 15, 1957; made available for printing October 8, 1957. This paper is now open for discussion; any discussions will be included in a subsequent issue.

W. F. WESTENDORP is with the General Electric Company, Schenectady, N. Y.

The author is indebted to M. A. Edwards and H. T. Boeker, of the General Electric X-Ray Department, for their prompt co-operation in trying out the various bias systems.

Fig. 1. One-million-volt electron beam generator for laboratory use

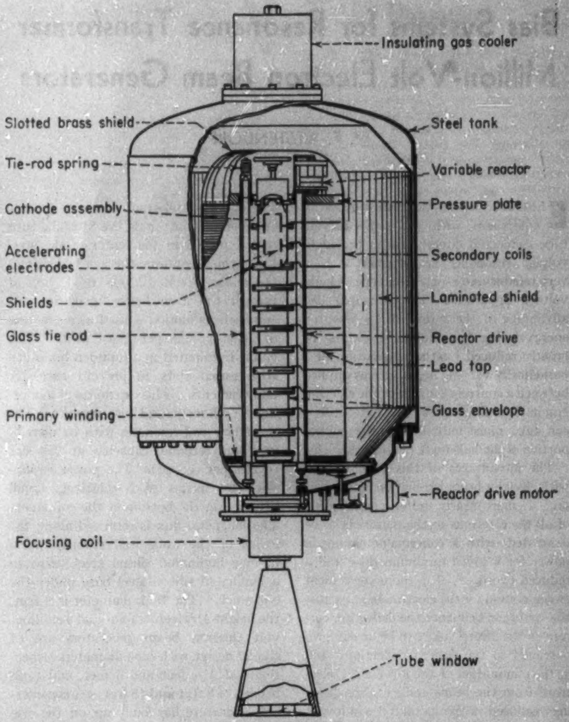


Fig. 2. Cross section of 1-million-volt electron beam machine with tank, transformer, and vacuum tube

sistor in the primary in order to maintain the current wave centered within the useful voltage half-wave.

Bias Independent of Operating Voltage

It is clear that in order to eliminate dependence on operating voltage no absolute reference should be used in generating the desired bias voltage wave. The iron core of the previous system saturates at a definite number of ampere turns, independent of operating voltage. If, instead, a reference level is used consisting of a voltage on a capacitor charged by means of a rectifier, and if it is used in generating the bias voltage, the reference level will vary with the alternating voltage and only the amplitude of the final bias wave will change, while its waveform remains constant.

Such a circuit is shown in Fig. 5. V_1 is an alternating voltage of the order of 8,000-volt peak current derived from the two top pancake coils of the resonance transformer coil stack. The operation is approximately as follows.

The input voltage V_1 of terminal B with respect to the conductor A-C is represented by the dashed sine wave with zero axis 0_1 in the lower diagram. This voltage charges the capacitor C by means of the rectifiers F_1 and F_2 to a potential V_3. During the time interval t_1, both rectifiers are conducting; since the forward drop of F_2 is much smaller than that of F_1, the output voltage V_4 from D to C will be very low and slightly positive. The excess of voltage V_1 of the sinusoidal wave over the capacitor voltage V_3 will be taken up by the drop across rectifier F_1. During the interval t_2, rectifier F_2 ceases to conduct and if it were not for the forward drop in F_1, the output voltage wave V_4 would coincide with the input wave V_1 if the zero axes of these waves are drawn respectively at 0_2 and 0_1. During time t_3, both rectifiers block current flow and the output voltage V_4 equals very nearly $R_2V_3/(R_1+R_2)$, the capacitor voltage V_3 being divided between R_1 and R_2.

Since the direct voltage V_3 is directly proportional to the input voltage V_1, a change in V_1 causes only a change in the amplitude scale and no change in the time scale of V_4.

This circuit was installed in the 1-million-volt X-ray machine of one of the company laboratories. It proved completely reliable. However, the current wave was not centered with respect to the voltage peak and came about 10 degrees late. Despite this phase error the X-ray output increased from 50 to 75 roentgens per minute at one meter by the introduction of bias.

Another disadvantage of this circuit and the previous one with the special peaking transformer was found in the X-ray department and consisted of a phase shift (lag) caused by an increase in tube current. In one test an increase

Table I. Effect of Bias for 1-Million-Volt and 3-Ma Operation

Conduction Angle, Degrees	Minimum Useful Voltage, Kilovolts	Average Useful Voltage, Kilovolts	Peak Current, Milliamps
180	0	637	6
90	707	900	12
80	766	922	13.5
70	819	940	15.4
60	866	935	18.0

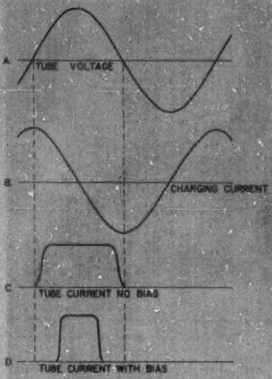


Fig. 3. Waveforms

A—Over-all tube voltage
B—Transformer secondary charging current
C—Tube current without bias
D—Tube current with bias

of tube current in a 1-million-volt X-ray outfit from 0.5 to 4 ma caused a phase lag of the current of 10 degrees. For X-ray work this is not serious since it will have only a minor effect on the quality of the radiation, but for cathode-ray work in which the beam is made to scan the material, it would require a readjustment of the scanning circuit and even with proper scanning the electron energy at the start of the sweep would be greater than that at the end. The explanation of the observed phase lag with increasing load current is that the voltage for the bias circuit is derived either from end turns or top coils, and since the induced voltage per turn in the top turns of the resonance transformer is approximately half that of the lower half of the coil stack, the inductive drop in these upper coils or turns causes a relatively larger phase shift with load than in the rest of the transformer. This phase shift in bias supply voltage finally results in an undesirable phase shift in tube current.

Bias Independent of Voltage and Tube Current

To eliminate the effect of load current on the phasing of the bias circuit the end turns had to be abandoned as a source of power. Instead, the phasing should take place with the total tube voltage as a reference. Consider the brass shield or "hat" on top of the transformer as a capacitor with respect to the grounded tank wall subjected to the full voltage. The current flowing from the winding to the hat leads the transformer voltage by 90 degrees. This current may be used to produce the desired bias voltage. It is approximately 30 ma rms in the 1,000,000-volt machines, 56 in the 2,000,000-volt machines, and 113 ma in the 4,000,000-volt machines.

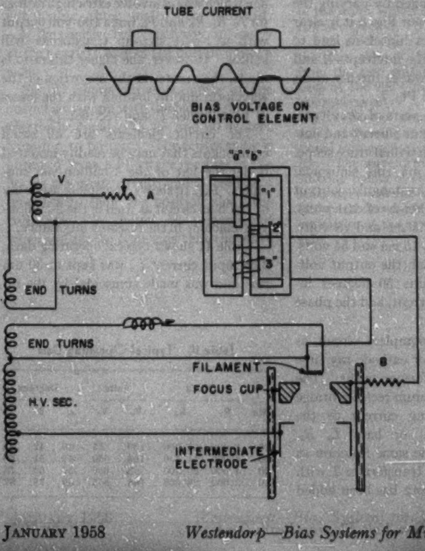


Fig. 4 (left). Bias from saturable core transformer

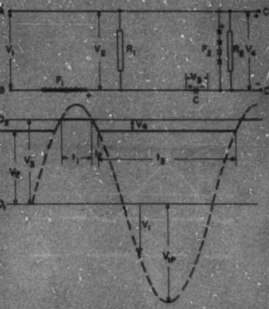


Fig. 5. Bias independent of operating voltage

Fig. 6 shows how this sinusoidal current of fundamental frequency (180 cps) may be converted into a rectangular asymmetrical voltage wave of twice fundamental frequency with the correct phase relation. The circuit shows a full-wave bridge rectifier with four branches F_1, a reactor L of such a value that the double frequency a-c ripple current through it is small compared to the direct current, which will establish itself through this reactor; a resistor R_1 to limit this direct current; a rectifier F_2 permitting current flow in the same direction as the output current from the F_1 bridge; a capacitor C which assumes a charge, the voltage of which is determined by the leak R_2. The capacitor is of such a value as to keep the double-frequency a-c ripple voltage across

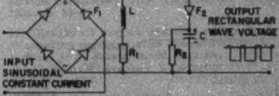

F₁
L
F₂
OUTPUT RECTANGULAR WAVE VOLTAGE
C
R₁
R₂
INPUT SINUSOIDAL CONSTANT CURRENT

Fig. 6. Circuit for bias independent of tube voltage and current

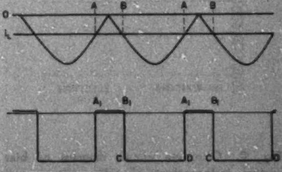
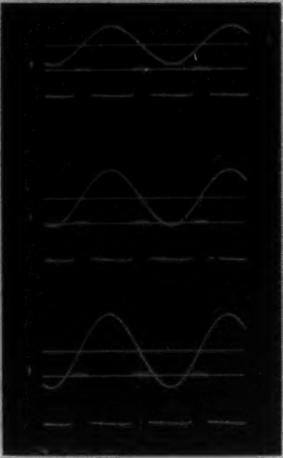


Fig. 7. Current and voltage waveforms



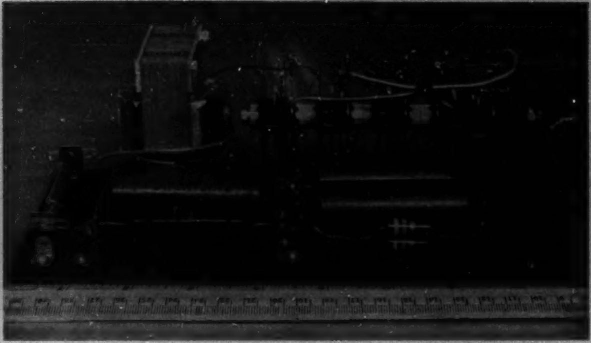


Fig. 8 (left). Sinusoidal input currents of 15, 22.5, and 30 ma, top to bottom, with rectangular output voltage waveforms

Fig. 10. Bias circuit assembly used in 2,000,000-volt machine in research laboratory

it small compared to the direct voltage it acquires. The operation is as follows: If the scalloped wave of Fig. 7 represents the available rectified sinusoidal current and the horizontal line i_L the constant direct current through L, then i_L exceeds the available rectified current during the intervals AB. This excess current of i_L over the sine wave will, however, circulate through the F_1 bridge with very low drop represented by the positive horizontal parts of the voltage drop curve of Fig. 7, labeled A_1B_1. During the intervals BA, however, the available rectified current represented by the scalloped wave exceeds the direct current i_L. The excess of current is now forced to flow through the rectifier-capacitor branch of the circuit, which has a voltage drop at least equal to the direct voltage on the capacitor. In Fig. 7 this shows up in the offset portions of the voltage wave CD. The lines DA_1 and B_1C are nearly perpendicular to the time axis. Charging currents of the distributed capacity and eddy current losses in L account for the slight deviation from the perpendicular position of these lines. The output voltage of the circuit may be increased in amplitude by increasing R_1 and R_2. The ratio of A_1B_1 to CD of Fig. 7 may be changed by varying the ratio of R_1 to R_2. From Fig. 6 it is clear that a lower R_1 and a higher R_2 lead to increased i_L, moving the intercepts A and B with the half-sine waves further apart and increasing A_1B_1 in Fig. 7.

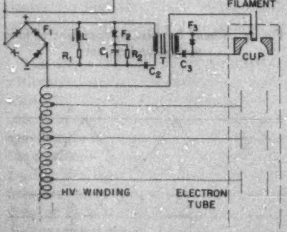


Fig. 9. Complete circuit diagram of bias system independent of tube voltage and current

Fig. 8 shows three pairs of simultaneous oscillograms of input current and output voltage taken with a dual-trace scope. The top pair is 30 ma rms sinusoidal input current and a rectangular output voltage with an amplitude of 101 volts. The middle pair is 22.5 ma and 78 volts, and the bottom pair is 15 ma and 54 volts. The angle during which the output voltage is positive remains 50 degrees independent of input current, and the phase remains correct.

Fig. 9 shows the complete circuit as applied to the X-ray or cathode-ray tube of a resonance transformer outfit. The input to the F_1 germanium rectifier bridge circuit is the charging current to the high-voltage terminal or hat. L, R_1, F_2, C, and R_2 have the same function as in Fig. 6. An output transformer T with a low loss Mumetal core has been added which has d-c blocking capacitors C_2 and C_3 in primary and secondary circuits, and a direct voltage component restoring rectifier F_3 consisting of a number of matched germanium diodes in series. The purpose of the transformer is to isolate the filament and focusing cup of the electron tube from the waveforming circuit and to provide a voltage step-up. If a 600-volt negative bias is needed at the tube, it will not be necessary to design the waveforming circuit for this output voltage which would involve extra high ratings for F_1, R_1, R_2, and F_2, but a 200-volt output with a 1-to-4 step-up transformer will suffice. However, the higher this ratio is made, the greater is the distortion of the waveform due to loading with the losses in transformer T, and rectifier F_3.

The circuit elements are all small components that may be readily mounted under the hat of the 1-million-volt machine. Fig. 10 shows a complete assembly of the bias circuit as used in the 2-million-volt machine in the research laboratory.

Table II shows typical operating data. The input current i_{in} was kept at 30 ma rms and was made sinusoidal by the in-

Table II. Typical Operating Data

i_{in}, Ma	Ohms		Volts		Degrees		
	R_1	R_2	V_c	V_e	α	β	γ
30	8,000	12,000	169	655	30	42	48
30	6,000	14,500	168	650	42	51	60
30	4,000	20,500	168	650	57	66	75
30	3,000	28,000	168	655	69	78	87

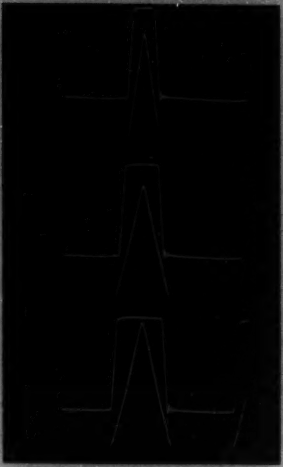


Fig. 11. Oscillograms of output voltage of complete circuit with duration at base of 43, 65, and 80 degrees, top to bottom. Spike is for time reference and shows input current zero

sertion of a series-resonant circuit with inductive and capacitive voltage drops of the order of 1,000 volts.

For the measurements R_1 and R_2 were decade resistors adjusted so as to maintain constant peak output voltage V_0 or negative bias for various conduction angles. Since the slopes of the switching intervals of the output voltage wave are not infinite, three angles were recorded: α at the positive top, β halfway up, and γ at the negative base of the output voltage wave.

Fig. 11 shows three examples of the output waveform with the base angle γ respectively, equal to 43, 65, and 80 degrees from top to bottom. The center spike is obtained from a full-wave rectifier in the 30 ma a-c input circuit, and is used as a reference to show the position along the time axis. It should be remembered when scaling the time axis that the wave is a 360-cycle wave and only every other positive bias pulse is utilized.

Fig. 12 shows the final electron current of the cathode-ray tube and the voltage as taken with a 1-million-volt machine, equipped with the latest bias circuit running at 1,000,000-volt and 3.0-ma average current.

Appendix

The values of the resistors R_1 and R_2 may be computed for the circuit of Fig. 6. With the assumptions that $\omega L >> R_1$ and $R_2 >> 1/(\omega C_2)$ it may be shown that the ratio R_1/R_2 is the following function of the conduction angle A:

$$R_1/R_2=(1/\pi^2)\\ [2(\pi-A)\tan 1/2(\pi-A)-(\pi-A)^2]$$

Also it may be shown that the d-c voltage V_c developed across C and the crest value i_m of the sinusoidal input current are related to R_1 and R_2 as follows:

$$R_1=(\pi-A)V_c/[i_m \cos 1/2(\pi-A)]$$

$$R_2=\pi V_c/\\ \{2i_m[\cos A/2-1/2(\pi-A)\sin A/2]\}$$

Computed values are within experimental error of observed values.

Second-order effects were computed, such as the effect of a ripple in the current through L, a ripple in the voltage across C, the effect of a resistance load across the output terminals and the effect of a capacitive load.

For brevity, only the qualitative results will be summarized. The result of the ripple current in a finite inductance L is a phase advance of the output voltage, which is inversely proportional to L.

Finite value of the capacitor C does not affect the timing of the output voltage wave but causes a ripple in the negative portion of the wave, which is inversely proportional to the capacity.

The effect of a resistance load is to slant the steep sides of the output wave outward from the zero voltage interval with a time base of the slant inversely proportional to the resistance.

The effect of a capacitive load is a time delay of the output voltage, which is proportional to the square root of the added capacity. The waveform is distorted so that the starting points of the steep portions remain stationary and the steep sides have become half-parabolas with the apex of maximum curvature at the starting point of the steep side.

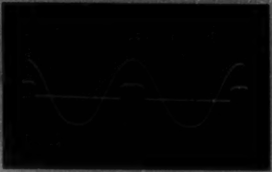


Fig. 12. Oscillograms of tube voltage and tube current for 1-million-volt and 75-degree conduction

References

1. A New Million-Volt X-Ray Outfit, E. E. Charlton, W. F. Westendorp, L. E. Dempster, George Hotaling. *Journal of Applied Physics*, New York, N. Y., vol. 10, no. 6, June 1939.

2. A Million-Volt X-Ray Unit, E. E. Charlton, W. F. Westendorp, L. E. Dempster, George Hotaling. *Radiology*, Syracuse, N. Y., vol. 35, no. 5, Nov. 1940, pp. 585–96.

3. A Portable Million-Volt X-Ray Outfit for Industrial Laboratories, E. E. Charlton, W. F. Westendorp. *General Electric Review*, Schenectady, N. Y., vol. 44, no. 12, Dec. 1941, pp. 652–61.

4. Two Million-Volt Mobile X-Ray Unit, E. E. Charlton, W. F. Westendorp. *Proceedings*, National Electronics Conference, Chicago, Ill., Oct. 1944.

5. Mobile Industrial X-Ray Unit, E. E. Charlton, W. F. Westendorp. *Electronics*, New York, N. Y., Dec. 1944.

6. An Oil-Immersed X-Ray Outfit for 500,000 Volts and an Oil-Immersed Multi-section X-Ray Tube, E. E. Charlton, George Hotaling, W. F. Westendorp, L. E. Dempster. *Radiology*, vol. 29, no. 3, pp. 329–40, Sec. p. 338.

7. The Dielectric Strength of Insulating Fluids, E. E. Charlton, F. S. Cooper. *General Electric Review*, vol. 40, 1937, p. 438.

8. Effects of Cathode Rays on Various Materials, E. J. Lawton. *Science*, Washington, D. C., vol. 112, 1950, p. 419 (abstract).

9. Initiation of Vinyl Polymerization by Means of High-Energy Electrons, J. V. Schmitz, E. J. Lawton. *Ibid.*, vol. 113, 1951, pp. 718–19.

10. Problems in Using High Voltage Electrons for Sterilization, W. D. Bellamy, E. J. Lawton. *Nucleonics*, New York, N. Y., vol. 12, 1954, pp. 54–57.

11. Effect of Initial Molecular Weight on Properties of Irradiated Polyethylene, E. J. Lawton, J. S. Balwit, A. M. Bueche. *Industrial and Engineering Chemistry*, Washington, D. C., vol. 46, 1954, p. 1703.

12. Effects of Atomic Radiation on High Polymers, K. H. Sun. *Modern Plastics*, New York, N. Y., Sept. 1954. (This article contains an extensive bibliography.)

The Examination of Electric Contacts by the Plastic Replica Method

H. W. HERMANCE
NONMEMBER AIEE

T. F. EGAN
NONMEMBER AIEE

ELECTROMECHANICAL switching devices, typified by the relay, are used by millions in the communications industry and have been developed to a high degree of perfection and specialization. Their reliability is such that only a few failures per billion contact operations are expected under telephone operating conditions. The use of twin contacts of noble metals, improved contact actuation, erosion control, and cleaner operating environments all have contributed to this perfection.

Nevertheless, trends in circuitry are such as to place still greater demands on these devices. Relays must operate faster on lower currents and with lighter contact pressures. The contacts must pass smaller currents with closer contact resistance margins. This is especially true where the aggregate resistance of long contact chains must be held to a low value. While relay design accomplishments have been notable, performance demands have more than kept pace with them, and thus the need for greater reliability still continues.

The vulnerable point in the relay is, of course, the contact. While noble metals such as palladium, platinum, or gold alloys have largely eliminated tarnish films, contacts are still subject to contamination from a variety of sources. Air-borne dusts, fumes, reactive gases, products of erosion, friction, and wear all may contribute to the impairment of contact function.

The design of contacts, the control of their operating environment, and the development of maintenance methods to minimize contamination effects have been continuing objectives in the Bell System. To meet these objectives, the engineer needs to know as much as possible about the nature of the contaminating substances and the way in which they behave on the contact. Some materials, for example, are fragile and poorly adherent. They crush easily to permit electrical closure and hence rarely give rise to "opens." Other materials may adhere tenaciously and are tough enough to resist the impacts of thousands of contact closures before their attrition eventually reestablishes electrical continuity. Arcing tends to burn away many organic materials, making a contact less susceptible to these contaminants. On the other hand, both arcing and friction in special cases may generate contaminating deposits.

Thus, the ultimate impairment may result from the complex interplay of impact, sliding friction, and arcing acting on a wide variety of gaseous, liquid, or solid contaminants present in the environment. Such effects can be understood only when critical methods of observation and diagnosis are used. Direct examination of contacts with various magnifying devices is a logical first step but this procedure necessarily has serious limitations, especially when the contacts are on equipment in service. Rarely can enough resolution be so obtained for positive identification. The contact surface provides a poor background for viewing the contaminant. It does not permit the use of transmitted light by which form, structure, and color may be determined. Chemical studies on the contact surface are almost out of the question. Occasionally, by extremely delicate and tedious micro-manipulative technique and with good luck, a deposit may be transferred to a microscope slide for further study. Even then, removal of the deposit from its original site deprives the observer of valuable diagnostic evidence. The relation of the deposit to the critical area of electrical closure is usually lost in this procedure and the interpretation of the part it may have had in causing failure is always open to question.

A technique of contact study which circumvents most of these difficulties has been in use for a number of years at the Bell Telephone Laboratories. The method consists of pressing the warmed contact into a clear thermoplastic material such as Vinylite. When cool, the resulting imprint reproduces faithfully all surface details of the contact in reverse relief. As an illustration of the detail obtainable by replication, Figs. 1(A) and (B) show photomicrographs of a polished, etched metal specimen together with a Vinylite replica of the same surface. The detail observable in the replica viewed by transmitted light appears to surpass that of the original specimen, photographed by reflected light. Most important, however, is the fact that solid deposits on the contact, almost without exception adhere to the plastic surface and thus are lifted off without appreciable shift of position relative to the contact topography. Thus, foreign material suspected of impairing contact operation is made available for microscopic study on an optically clear and relatively inert base to which various chemical tests may be applied.

Typical replicas of relay contacts are shown in Fig. 2. The area involved in closure and electrical transmission on a contact pair can be seen on the replica as texture changes produced by impact, sliding friction, and electrical erosion. The manner in which a foreign particle or deposit is related to this area becomes apparent under the microscope.

Replica sampling then, is a simple procedure which makes it possible to gather and to preserve evidence of contact contamination, wear, or erosion on equipment in service and to transmit this evidence, completely intact, from distant locations to a central laboratory for study and correlation. The replicas may be projected on a screen for discussion or photomicrographs may be made for reporting. The replica may then be easily filed away for future reference and comparisons.

Naturally, there are certain limitations to the replica method. Oily films or materials which would melt or decompose or dissolve in the plastic at the softening temperature, 135 to 145 C (degrees centigrade) for about 30 seconds would not be satisfactorily studied in this way. Adherent tarnish films may not transfer, although silver sulfide is often found to lift off onto Vinylite. Some materials, particularly masses of fine particles such as erosion powder, may undergo slight deformation under the flattening force necessary to make the replica. In the authors' experience, however, this effect rarely has been great enough to prevent interpretation.

When a loose, powdery deposit is present in considerable depth on the contact surface there will be a limit to the quantity which can be lifted off because of saturation of the tacky surface.

Paper 57-1095, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 7–11, 1957. Manuscript submitted May 21, 1957; made available for printing September 6, 1957.

H. W. HERMANCE and T. F. EGAN are with Bell Telephone Laboratories, Inc., Murray Hill, N. J.

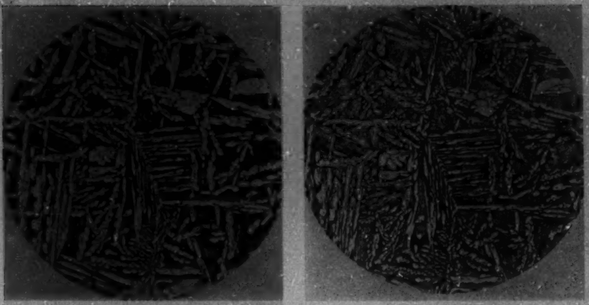


Fig. 1. Example of replication (200×)

A—Original specimen of aluminum bronze by reflected light
B—Replica of aluminum bronze surface by transmitted light

Several replicas may then be necessary for complete removal.

Vinylite *VS-1310* is generally satisfactory as a thermoplastic material. Vinylite sheet is easily cut or punched to form the small blanks used on microcontacts. It also has a reasonably low softening point, good impressions being obtained at 135 to 145 C. The softened Vinylite is tacky and most deposits on the contact adhere to it tenaciously.

Vinylite has high clarity for microscopic examination and is inert enough chemically to withstand strong reagents. Concentrated sulfuric acid droplets, for example, can be applied to the replica for 5–10-minute periods without appreciable attack. Exposure to chlorine or *aqua regia* fumes for 15-minute periods results in little damage. Naturally, organic solvents must be limited to those in which the plastic does not soften.

It has been determined by microscopic observations, X-ray diffraction studies, and delicate resistance measurements that, when properly performed, replication with Vinylite transfers no detectable contaminations to the contact. In fact, the hot plastic transfer has been found to be a highly effective method for cleaning contacts used in laboratroy studies.

In making replicas of small contacts, 0.015- to 0.030-inch sheet is used. This is cut or punched into blanks of size and shape suited to the particular contact arrangement. The blanks are rinsed in water containing a wetting agent (Aerosol *OT*), then in 80% alcohol and blotted dry with lens tissue. They are individually inspected under a 40x stereoscopic microscope and those having bad scratches or imbedded material are discarded. Any remaining loose dust particles are removed under the microscope with a small glass suction tube. The clean blank is placed in the cavity of a

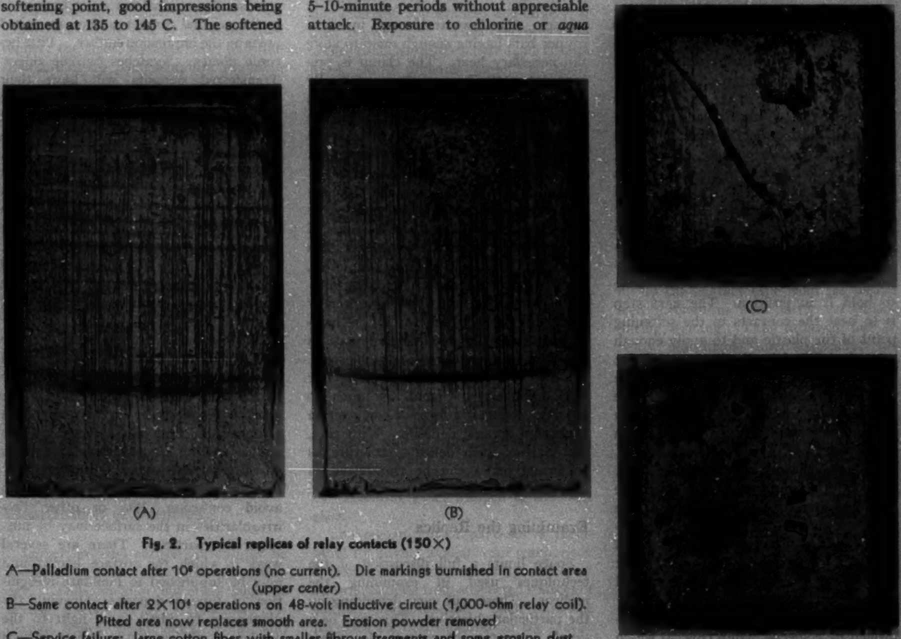


Fig. 2. Typical replicas of relay contacts (150×)

A—Palladium contact after 10^6 operations (no current). Die markings burnished in contact area (upper center)
B—Same contact after 2×10^4 operations on 48-volt inductive circuit (1,000-ohm relay coil). Pitted area now replaces smooth area. Erosion powder removed
C—Service failure: large cotton fiber with smaller fibrous fragments and some erosion dust
D—Service failure: phenol fiber fragments predominating

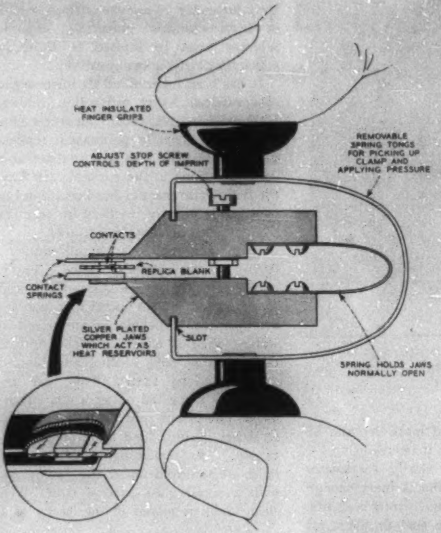

HEAT INSULATED FINGER GRIPS
ADJUST STOP SCREW CONTROLS DEPTH OF IMPRINT
REMOVABLE SPRING TONGS FOR PICKING UP CLAMP AND APPLYING PRESSURE
CONTACTS
REPLICA BLANK
CONTACT SPRINGS
SILVER PLATED COPPER JAWS WHICH ACT AS HEAT RESERVOIRS
SLOT
SPRING HOLDS JAWS NORMALLY OPEN

Fig. 3 (left). Details of heating clamp

Fig. 4 (right). Insertion of Vinylite disk between contacts

special microscope slide. The cavity is closed by a rectangular cover glass hinged at one end by pressure-sensitive tape. The free end of the cover glass is held down by a small tab of the tape. In this way a supply of clean blanks can be prepared ahead for field studies and stored in standard indexed plastic slide boxes. The replica is returned to the slide as soon as it is made, reducing the hazard of contamination by air-borne dusts.

To make the replica, the blank is introduced between the opened contacts without smearing their surfaces. The contacts are then gently closed on the plastic to hold it in position. The next step is to heat the contacts to the softening point of the plastic and to apply enough pressure to produce the imprint. These operations will vary somewhat with the contact actuation and arrangement. The U-type telephone relay serves as an example of an uncomplicated design which lends itself well to replica sampling. Here disk-shaped blanks are used. Initially, the contacts should be in the closed position. If normally open, the armature may be held operated to provide closure. The contacts are then opened for insertion of the replica blank by means of a tapered glass probe applied between the springs.

Heating of the relay contacts is accomplished by a special clamp (Fig. 3) which is closed on the backs of the contact springs. The jaws are formed by short copper bars having enough mass to store the necessary heat. The clamp is preheated to 140 C on a thermostatically controlled heater block. It is applied to the contact springs by special tongs with heat-insulated finger grips. The tongs engage in slots milled in the copper bars. Details are shown in Figs. 4, 5, and 6.

After the impression is made, the contact must be allowed to cool before the plastic disk is removed. The depth of the impression should be no greater than necessary to include the area of interest. Deep impressions introduce lighting difficulties under the microscope, especially when projected or photographed. Chemical treatments and micromanipulations are also rendered more troublesome. Heating should be continued no longer than necessary to avoid, as much as possible, changes in form or composition of foreign organic deposits. With care, materials of such delicate structure as epidermal scales can be transferred in identifiable condition.

Examining the Replica

Space limits a full discussion of the techniques used in examining the replicas. These will vary depending on the particular objective. This may be a simple inspection of contact cleanliness as part of a routine quality control. On the other hand, it may be a more fundamental study of contact contamination requiring specific identification of all foreign materials. Replicas also may be used to record contact wear or erosion.

For examining the replica, a microscope with substage condenser and provision for polarized light will be needed. In addition to viewing in transmitted light, means for illuminating the replica with angular incident light should be provided.

The replica disk will contain an impression on each side, corresponding to the two mating contacts. The impression reverses the original texture. Low spots in the original contact surface are raised areas on the impression surface. Pitts become domes. Scratches become ridges. Transferred deposits will have their original exposed surfaces embedded in the plastic and the surfaces originally adjacent to the contact metal will now be the exposed portion. Under the microscope, some of the detail of the lower impression can be observed by focusing down through the upper one. This procedure is useful when one wishes to compare corresponding points on the two contacts, such as the areas of slide and electrical closure. It also permits one to view the originally exposed surface of a deposit, now imbedded in the plastic.

The surface texture details of the replica are best viewed by transmitted light, and they may be intensified by partially blocking the light entering the microscope condenser so that the replica is illuminated from one side only. When one is looking for transferred particles on the replica, however, it is often desirable to subdue the texture patterns to avoid confusion. Pits or other tiny irregularities in the surface may be mistaken for particles. There are several ways in which the surface background may be suppressed to facilitate recognition of foreign deposits. Opening the diaphragm to admit more light to the condenser will cause the lighter surface features to disappear, while all but the

Fig. 5 (left). Heating contact springs with clamp

Fig. 6 (right). Returning replica to cavity in slide

most transparent particles stand out in strong contrast. Diffusion of the light by interposing a ground glass plate between the condenser and the replica also reduces the texture contrast and permits easier recognition of particles and films. Mounting the replica in Canada balsam or other media of similar refractive index eliminates texture and at the same time renders the color, transparency, and fine structure of the particle more observable (Fig. 7). Polarized light too is often useful in bringing out differentiating structures and in making small mineral or fibrous fragments detectable. Strains in the plastic will often produce patterns in polarized light, but these are easily distinguished from the sharp boundaries of optically active particles. Fig. 8 compares the appearance in ordinary and polarized light of phenol fiber particles and a calcium carbonate paint pigment.

Viewing by incident light often provides valuable clues in the identification of particles. The luster and color by reflected light and the surface texture of the particle when the light strikes at a low angle are all helpful. The slide may be placed on a variety of backgrounds selected to provide the best contrast for viewing the particle. A flat white and a light-absorbing (optical) black background are the most useful.

Identification of Particles

Larger, isolated particles often may be identified by their characteristic morphology. Thus, cellular materials such as textile fibers, woody fragments, epidermal scales, plant hairs, seeds, and pollen have structural features and characteristic outlines which make their recognition fairly certain by an experienced microscopist.

Natural minerals such as quartz, feldspar, calcite, asbestos, etc., also have characteristic appearances due largely to their crystallinity. Thus fracture, even of tiny particles, often occurs along crystal faces. Most minerals present in dusts are optically active and polarized light may reveal particles in the replica too small to be noticed in ordinary light. The sharp extinction of most mineral crystals when the stage or slide is rotated in polarized light contrasts with the indefinite behavior of most cellular materials.

Another particle present universally in the dusts over bituminous coal-burning areas is "fly ash." This is a glass, formed by fusion of coal ash dust in hot combustion gases. It is usually colored, probably by iron and other metals, and ranges from light yellow to deep amber. Inclusions of unburned carbon, minute gas bubbles, and unfused ash are often present and serve to identify this material. The shape of the particles is often distinctive. They are frequently very thin and curved with angular, broken outlines. When so found, they are derived from bubble-like spheres formed by expanding gases trapped in the fluid glass.

The recognition of particles is facilitated by having available for comparison a collection of permanently mounted known particles from suspected sources of contact contamination. Such reference particles are best mounted in the plastic used for making the replicas. Comparison is then made in the same medium.

Contact contaminants are not always easily identifiable isolated particles. Often they are completely amorphous substances without characterizing structure or optical behavior. They may consist of impacted or smeared aggregates made up of several materials. A dark granular smear, for example, was found on replicas of contacts involved in service failures in a new central office (Fig. 9). Although devoid of clear-cut identifying characteristics, this material could be recognized repeatedly by its dark heterogeneous and smeary nature. By systematic replica sampling, this particular complex contaminant was traced back to the relay assembly line. Here it was found that gummed oil mixed with finely divided metal and atmospheric dusts was picked up by the contact tape from the welder feed mechanism and partly carbonized during welding. Thus although not directly identifiable microscopically, the nature of a contaminating material may be revealed when its source is systematically traced.

Where microscopically unidentifiable materials are encountered, chemical treatments may aid the investigator. Under such treatment, the behavior of some materials may be specific enough for identification. With others, chemical behavior at least may narrow classification of the material.

Because interest centers on specific particles or aggregates, chemical reagents should be applied to the replica so as to confine the reaction products to the sites occupied by the original particles. The simplest and most satisfactory way

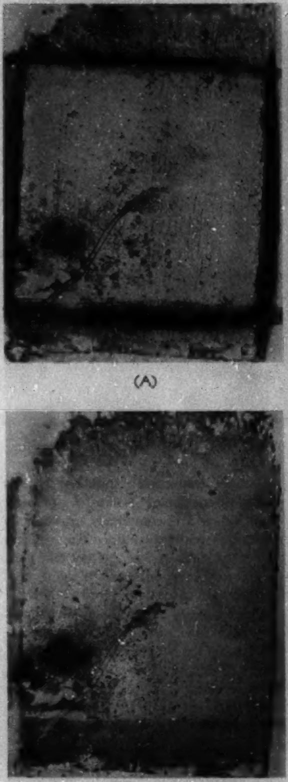

(A)
(B)

Fig. 7. Replica of silver contact (150×)

A—In air
B—Immersed, showing silver sulfide film

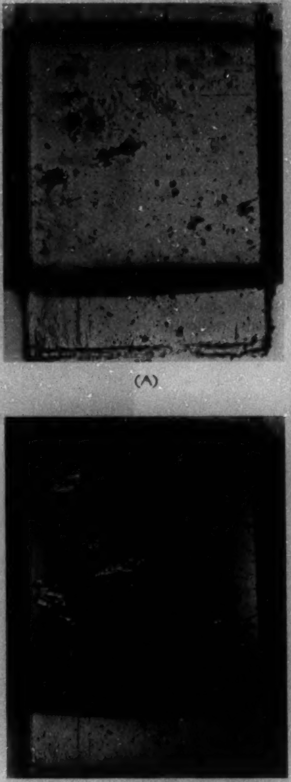

(A)
(B)

Fig. 8. Replica with fiber fragments (150×)

A—In ordinary light
B—Between crossed Nicols

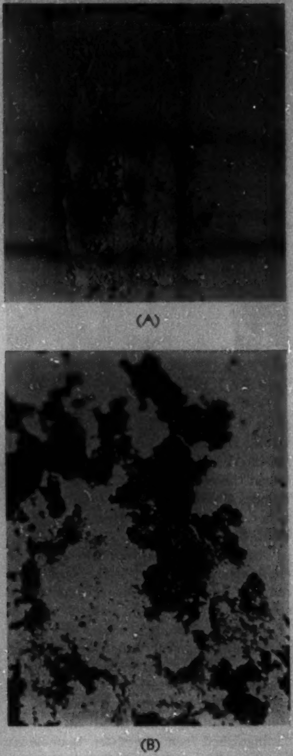

(A)
(B)

Fig. 9. Palladium contact contaminated in welding operation by oxidized and partly carbonized oil and finely divided palladium

A—View of contact
B—Outlined area in A at higher magnification showing particles of palladium in the film

of doing this is to expose the replica to a reactive gas or vapor. When this can be done, the replica is placed in a suitable fuming chamber and there is no need for delicate manipulations on minute areas (Fig. 10). Thus acid vapors such as hydrochloric, nitric, or *aqua regia* may be employed to dissolve metal, metal oxide, or certain mineral particles (Fig. 11). In place of the particle, crystals of its salt or droplets of its concentrated solution will form. The color of the solution droplet or the color, form, and solubility of the crystalline deposit may be enough to identify the metal. If not, other vapor treatments designed to further the identification may follow. For example, exposure to ammonia may produce characteristic color changes or precipitates in the droplets.

Iodine vapor in low concentrations is taken up by oils, fats, and resins, imparting a strong orange-yellow color. Epidermal scales, which contain fatty matter, also stain yellow with iodine as do other oil-containing plant or animal tissues. Starch-containing materials in the presence of moisture form the blue addition compound with iodine.

Solvents too may be applied to the replica as vapors. The solution process then is held under better control and may be watched under the microscope. In taking up solvent, amorphous particles usually swell, then gradually lose their form, eventually becoming droplets. Particles of many soluble salts will dissolve when exposed over water to form tiny pools of solution. Naturally, solvents which strongly affect the replica plastic must be avoided.

Where vapor treatments cannot be devised, reagents must be added to the particles as minute droplets. Localized droplets may be applied from pipettes with very fine capillary tips while the replica is observed under the microscope. With practice, it is possible to control the placement and size of the droplet so as to involve only an area of the replica in which the particle is located. The

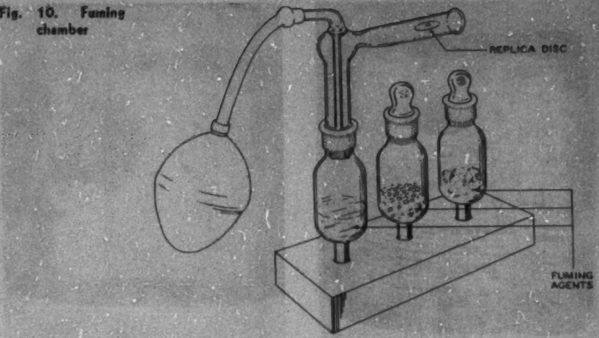


Fig. 10. Fuming chamber

reagent is removed by a reversal of the pipetting process and the area may be flushed, if desired, with several wash droplets manipulated in the same manner. Biological and chemical staining, carried out in this manner, are particularly useful.

The variety of microchemical tests which can be carried out on replicas is limited only by the patience and ingenuity of the investigator. Almost every metal ion can be identified by conventional reactions applied to the solutions obtained by a preliminary acid fuming. The sensitive color and crystal tests devised by Feigl, Chamot, and others are readily applicable. Special tests usually can be devised. Thus the addition of a silver nitrate droplet will produce needle-like clusters of silver crystals with metals above silver in the electromotive series. A droplet of glycerin containing hydrochloric acid will trap and hold bubbles of carbon dioxide evolved from a carbonate long enough to permit detection. Chromic acid in concentrated sulfuric acid will prove the organic nature of a particle by slow reduction of the orange chromic acid to the green trivalent chromium and, at the same time, often forming carbon dioxide bubbles.

Table I lists the materials identified by the authors on about 2,000 replicas of telephone relay contacts. These contacts were in a wide variety of environments ranging from extremely dirty to extremely clean conditions. Many of the replicas were of contacts known to have failed in service.

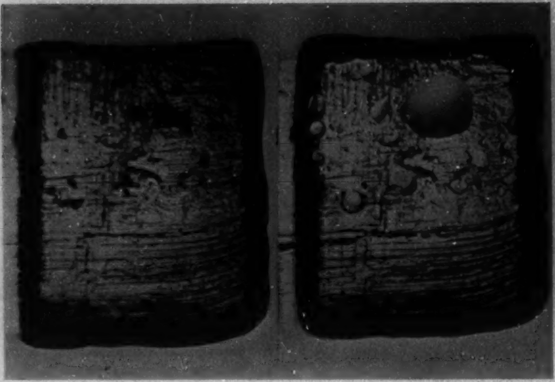
Application of the Replica Technique

The replica method has had many applications in the telephone field. One was a systematic study of relay contacts which occasionally develop high resistance. Most telephone relays have twin contacts so that the likelihood of a circuit failure from a transitory contamination of both contacts is remote. Tenacious particles or deposits which resist attrition, however, may put one of the twins out of service for several hundred or even thousands of operations. Then the relay will be vulnerable to even fragile air-borne dusts. This is the condition under which most service failures occur. To understand better the nature of the persistent type of contact contamination, replicas were made of each open contact to which a circuit failure could be traced. This was done in a number of central offices in several cities. The results indicated that over 90% of the opens were caused by dusts of a fibrous character. Of the textile fibers, cotton predominated. These fibers were traced largely to clothing lint shed by workers during manufacture, installation, and operation. Packing materials were also responsible for some fibers, particularly from the poorer grades of wood pulp paper. A particularly troublesome material was the phenol fiber fragments produced during punching or milling of the spoolheads, cards, and insulators. Fibrous fragments under the impact of closing contacts are sufficiently plastic to lock into any surface irregularities and are not likely to be jarred off by subsequent relay operation. Even compressed-air cleaning of the equipment may not remove such impacted particles. Contacts which arc are less susceptible to fibrous dusts. Replicas of the few open arcing contacts show large particles which evidently have deposited directly across the area of electrical closure. Otherwise, the arc tends to char and to blast away any organic dusts deposited peripherally.

Table I. Particles and Deposits Found on Contact Replicas

Material	Size Range, Microns*	Source
Fibrous:		
Cotton	100	Clothing and insulation
Wood	100	Clothing
Silk, rayon, and acetate	100	Clothing and insulation
Nylon and Orlon	100	Clothing
Phenol fiber	25–100	Machined insulation
Wood pulp	50–100	Paper
Plant hairs	50–100	Seeds, stems, etc.
Other cellular:		
Epidermal scale	25–50	Shed from skin
Insect scales and parts	50–100	
Pollen	25–50	
Woody tissue	25–50	Disintegrated plants, etc.
Combustion products:		
Cinder glass	1–25	Soft coal ash
Anthracite ash	1–15	
Coal	1–15	
Soot flocks		Coal, oil, and gas
Minerals:		
Quartz	1–25	Soil and rock erosion
"Earthy" aggregates		Soil and rock erosion
Plaster	1–25	Building activities
Asbestos	25–50	Insulation
Metals and corrosion products:		
Copper		Frictional contamination
Fe_2O_3 and Fe_3O_4	1–15	Iron rust and mill scale
Lead-tin alloys		Frictional contamination
Silver sulfide		Corrosion on contact
Nickel sulfate		Corrosion products on contact spring
Zinc sulfate		Corrosion products on contact spring
Organic deposits:		
Wire enamel	10–50	
Oxidized oils		Lubricants, paint
Rosin	1–10	Flux fumes
"Polymer"		Smeary brownish deposit on platinum metals from adsorbed organic vapors

* Size range most frequently found.



(A) (B)

Fig. 11. Contact replica showing palladium fragments

A—Before fuming
B—After exposure to aqua regia vapors. Droplets of palladium chloride form on sites of original palladium particles

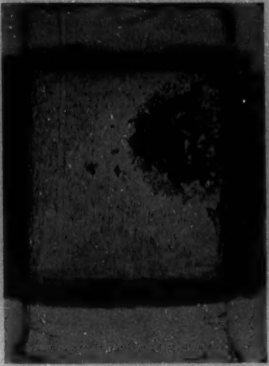


Fig. 12. Polymerized organic deposit on palladium contact (150×)

The observed importance of dusts produced by switchroom and manufacturing activities has tended to shift the emphasis somewhat away from complete filtration of the outside air supply as a cure-all for contact trouble. More attention is being directed toward control of dusts produced within the switching areas and during manufacture. New designs include better dust-protection enclosures. Improved housekeeping practices, limitation of personnel activities in switchrooms, better management of air flow patterns in ventilation, all have helped to reduce the hazards of fibrous dusts.

Certain contaminants responsible for relay failures have been traced back to manufacture. Among these, phenol fiber dust and smears of gummed, carbonized oil have been mentioned earlier. Materials such as paper, wood, epidermis, rosin, insulating enamel, iron rust, fly ash and cinders, and foreign metals such as copper, brass, solder, etc., and their oxides also have been identified on the contacts of the finished relay. Such revelations have brought about changes in manufacturing practices, with the replica sampling method serving as the basis for quality control.

A peculiar and completely unexpected contaminant was discovered several years ago on replicas of nonarcing palladium contacts taken in the field. On the replica there was invariably found a brownish, amorphous, transluscent material in a zone immediately surrounding the area of contact closure; see Fig. 12. Viewed on the contact itself, this deposit might easily have been confused with wear and erosion products. This probably accounts for its not having been discovered earlier, since palladium has been long in use. On the replica, however, in transmitted light, the unique nature of the material is readily apparent.

This disclosure led to subsequent laboratory studies in which it was shown that when palladium surfaces are rubbed together in the presence of traces of organic vapors there is formed a strongly cross-linked substance of high molecular weight which deposits around the area of slide. The phenomenon is specific for the platinum-group metals and occurs with a great variety of organic vapors. The traces of vapor emanating from the phenolic resins and plasticizers in the relay structure close to the contacts provide a sufficient vapor source, particularly when they are confined by enclosure of the relay. It is believed that the vapors are first chemisorbed on palladium, then polymerized. The part played by sliding friction is still under study. It may consist simply in a renewal of the surface by wiping off the polymer, it may involve some sort of frictional activation of the metal, or it may provide thermal effects necessary for completion of the polymerization.

ERRATA

"Electrical Units and Dimensions" by Leo Young, published in AIEE Transactions, part I, 1956 (January 1957 section), pages 767–71.

The symbol shown as E_0 which appears below equation 13 on page 768 should be ϵ_0.

"Transient Analysis of Second-Order Flip-Flops" by L. M. Vallese, published in Communication and Electronics, May 1957, pages 161–66.

Reference 7 incorrectly indicates that the article, "Stability Analysis of a Basic Transistor Switching Circuit" by T. R. Bashkow, appeared in the 1954 *Proceedings* of the National Electronics Conference, Chicago, Ill. The article may be found in the foregoing *Proceedings* for 1953, volume 9, page 748.